INTEGRATIVE ASSESSMENT OF SYSTEMATIC GENE EXPRESSION VARIATION IN RESPONSE TO OSMOTIC SHOCK AND ENVIRONMENTAL TOXICANTS

by

THOMAS HEYWARD HAMPTON

A thesis submitted to the University of Birmingham for the degree of DOCTOR OF PHILOSOPHY

School of Biosciences

University of Birmingham

July 2017

Abstract

This thesis applies integrative and systemic approaches to gene expression experiments measuring responses to environmental stress. Methods were developed to identify systematic differences in response strength, functional pathway activation, and gene regulatory network structure. Results in three wild killifish populations revealed high population variability at the level of individual genes, consistent with the killifish's genetic diversity and ability to adapt rapidly to anthropogenic pollution. Despite gene level diversity, modular network structures, patterns of pathway activation, and patterns of gene expression canalization were conserved in the three populations, demonstrating that gene regulatory networks are preserved by selective processes and may constrain killifish adaptation. The presence of arsenic during killifish acclimation to osmotic shock systematically reduced the magnitude of gene expression responses, and reduced coordination between genes that respond to osmotic shock. Results in the water flea suggested that cadmium tolerance is associated with systematically larger gene expression responses to cadmium stress, and greater network coordination among genes that respond to cadmium. In summary, environmentally responsive gene regulatory networks 1) shape the efficacy of biotic and abiotic stress responses, 2) are targeted by toxic effects, and 3) are shaped by selective forces.

Dedication

This thesis is dedicated to Bruce A. Stanton, who suggested I get a PhD, paid my salary during the process, and has been a supporter of my scientific advancement for almost ten years. Bruce is an outstanding mentor, but also a fine role model for anyone in science. He is curious, principled, sets high standards for himself, and is always willing to leave familiar scientific territory behind, learn whole new disciplines and collaborate with new people.

Acknowledgements

First, I want to acknowledge my thesis committee: John Colbourne, PhD, Joe Shaw, PhD, and Shan He, PhD. Thanks to John for doing the work to set this unique distance learning opportunity for me, providing funding and guidance in environmental genomics. Thanks to Joe for sage advice in environmental toxicology, and encyclopedic knowledge of killifish and Daphnia. Thanks to Shan He for participating in three years of meetings and providing insights into network biology.

Second, I want to collaborators on the three chapters, who will be co-authors of the research papers associated with them: Dörthe Becker, for advice and support related to the *Daphnia* project, Celia Y. Chen, for field work associated with the *Daphnia* and killifish projects, Dagmar Frisch, for advice and support related to the *Daphnia* project, Stephen P. Glaholt, for field work, animal husbandry and laboratory work related to the *Daphnia* and kiilifish projects, Craig Jackson for raw data and annotations related to *Daphnia* and killifish projects, Dawoon Jung for field work and animal husbandry related to the killifish project, Barbara Pietrzak for advice and support related to the *Daphnia* project, and Marcin W Wojewodzic for advice and support related to the *Daphnia* project.

Third, thanks to my wife Katja Koeppen, PhD and Bruce Stanton, PhD for being shadow members of my thesis committee, reading drafts, and telling me they didn't understand what I just said.

Fourth, thanks to all my scientific collaborators on a variety of projects, who have helped teach me the craft of scientific research, especially Karen Wetterhahn PhD, Joshua Hamilton PhD, Marc Ernstoff, MD, George O'Toole PhD and Andrew Whitehead PhD.

Finally, thanks to all the members of the Stanton Lab at the Geisel School of Medicine at Dartmouth, who have been very supportive. Special thanks to Britton Goodale, PhD for insights regarding arsenic and environmental toxicology, as well as Roxanna Barnaby, Zhongyou Li , Erica Rayack, Amanda Liefeld, and Victoria Holden, PhD.

Figures and Tables ..... **i**

Figures ..... **i**

Tables ..... **viii**

1. General Introduction **1**

What this Thesis Found **4**

Islands of Constraint Amid a Sea of Diversity 4

Arsenic as a Network Toxicant 5

Networks Facilitate Cadmium Tolerance 5

Methods **6**

Magnitude of Gene Expression Response 6

Coordination of Network Response 7

Pathway Activation 9

Tightly Controlled Genes (Gene Expression Canalization) 12

Background **13**

Killifish and Salinity Acclimation 13

Daphnia and Cadmium Adaptation 16

Cadmium Toxicity 17

Arsenic Toxicity 19

2. Conserved Gene Networks Facilitate the Osmotic Shock Response in Killifish 22

Introduction 24

Materials and Methods 27

Killifish Populations and Exposures 27

RNA Isolation 28

RNA Quantitation by Microarray 28

Gene Expression Data Analysis 29

Gene Set Activation Analysis 29

Extramural Network Cohesion 30

Gene Canalization 30

Results 31

High Gene Expression Variability Within and Among Killifish Populations 31

Known Osmoregulation Responses were Shared by Populations 37

Different KEGG Pathways were Activated at 1 h and 24 h 38

Networks of Salinity Response Genes were Loosely Coupled 40

Salinity Response Genes were Highly Canalized 42

Highly Canalized Genes Clustered by Geography 44

Discussion 45

Amid Population Variability, Familiar Responses are Conserved 45

Hypo-osmotic Shock Alters Cellular Resource Management 48

Modular Networks Facilitate Acclimation 49

Tightly Controlled Genes Define Geographic Clades 49

Conclusions 49

3. Low Dose Arsenic Reduces Gene Regulatory Network Connectivity and Impairs Phenotypic Plasticity in Killifish 50

Introduction 51

Materials and Methods 53

Killifish, Exposures, Tissue Collection 53

RNA Isolation, Hybridization and Normalization 54

Identifying Genes That Respond to Hypo-osmotic Shock 54

Identifying Genes That Respond to Hyper-osmotic Shock 55

Quantifying the Impact of Arsenic on the Magnitude of Osmotic Shock Responses 55

Quantifying the Impact of Arsenic on Gene Co-regulation 56

Quantifying the Impact of Arsenic on Pathway Activation 56

Results and Discussion 57

Arsenic Diminishes Gene Expression Response to Osmotic Shock 57

Arsenic Reduces Co-expression of Genes Responding to Osmotic Shock 60

Transcription Factors Significantly Induced by Hypo-Osmotic Shock 64

Arsenic Reduces Transcription Factor Co-regulatory Relationships at 1 h 66

Co-regulation loss at 24 h Associated with Reduced Response 68

Arsenic Interactions Inhibit Pathway Responses at 1 h 69

4. **Enhanced Gene Expression Response to Cadmium Stress May Facilitate Cadmium Tolerance in *D. pulicaria* 72**

Introduction 73

Materials and Methods 75

Animals 75

Tolerance Assessment 75

Toxicant Exposures 76

RNA Isolation and Extraction 76

Microarry Processing 76

Gene Expression 78

Identifying Differentially Expressed Genes 78

Identifying Systematic Differences in Response Magnitude 78

Pathway Gene Set Enrichment and Activation 78

Gene Regulatory Network Effects 79

Statistics 79

Results and Discussion 80

Sensitive and Tolerant Clones Differ at Baseline 80

Glutathione Metabolism Pathway Repressed at Baseline in Tolerant Clone 81

Cd-Sensitive and Cd-Tolerant Clones Share Many Pathway Responses to Cadmium 82

Tolerant Clone Activates Taurine and Represses Ribosome Pathway 84

Tolerant and Sensitive Clones Correlated in Cadmium Expression Response 85

Tolerant Response to Cadmium is Systematically Larger 88

Enhanced Response in Tolerant Clone is Cadmium-specific 89

Co-regulation of cadmium-sensitive genes is higher in tolerant clone 90

Supplemental Section 93

5. Synopsis 98

Summary of Results ..... Error! Bookmark not defined.

Selective Pressures Shape Gene Expression Response Networks 98

Cadmium Tolerance is Associated with Coordinated, Intense Gene Expression Response 99

Arsenic Degrades Gene Expression Responses and Networks in Killifish 99

Predictions Based on the General Model **100**

Experiments to Test Model Predictions **101**

Applying the Methods Developed in This Thesis to Other Research Domains **102**

Conclusions **102**

References **103**

Figures and Tables

Figures

Figure 1-1 Idealized reaction norms for two individuals, “red” and “blue”, responding to increased temperature. Red responds more in both changes in vasoconstriction and blood flow, as indicated by the larger reaction norms**.** 6

Figure 1-2 Simulated correlations (co-expression) between genes, scatterplot representation Gene 3, Gene 2 correlates with Gene 4, illustrating the concept of gene co-expression...8

Figure 1-3 Simulated correlations (co-expression) between genes, network representation: Gene 1 and Gene 3, and between Gene 2 and Gene 4, represented as a network diagram. 9

Figure 1-4 Idealized annotated signaling cascade. Gene 1 induces Gene 2, shown by arrows, overlaid with hypothetical gene responses to experimental conditions. Red is induction; green is repression, illustrating a discordance between observed gene expression responses and previously curated gene regulatory relationships. 10

Figure 1-1-5 Simulated gene expression data illustrating genes that are tightly regulated (canalized).....12

Figure 1-6 Major ion and water flows in fish inhabiting fresh water and seawater environments(Evans 2008). 14

Figure 1-7 A framework relating the ability to acclimate to osmotic shock to gene regulatory network structure. From: Shaw, J. R., Hampton, T. H., King, B. L., Whitehead, A., Galvez, F., Gross, R. H., et al. (2014). Natural selection canalizes expression variation

of environmentally induced plasticity-enabling genes. *Molecular Biology and Evolution*, *31*(11), 3002–3015. Reproduced by permission of William Scavone (http://www.kestrelstudio.com/). 15

Figure 1-8 Cyclic Parthenogenesis in *Daphnia* : Adapted from: Shaw, J. R., Pfrender, M. E., Eads, B. D., Klaper, R., Callaghan, A., Sibly, R. M., et al. (2008). *Daphnia* as an emerging model for toxicological genomics. In *Comparative Toxicogenomics* (Vol. 2, pp. 165–328). Elsevier. 16

Figure 1-9 Cadmium coordinated by four cysteine residues in an idealized metallothionein(Rubino 2015). 17

Figure 1-10 Methylation of arsenite (Hughes *et al.* 2011).....19

Figure 2-1 Gene expression variation distinguishes populations. Principal components analysis of log transformed, normalized gene expression values for 36 saltwater-acclimated killifish from King's Creek, VA (yellow), Northeast Creek, ME (blue) and Horseshoe Creek, ME (red) following exposure to fresh water for 0, 1, or 24 hours. X axis shows loadings associated with the first principal component (PC1) which accounts for 25% of the variability of gene expression; the second component (y axis) accounts for 19% of variability. 32

Figure 2-2 Fish from the same population are significantly more alike. Pairwise Euclidean distance measures for fish from the same population, but different hours of exposure to fresh water (left) or fish exposed to fresh water conditions for the same time, but not from the same population. Boxplots represent 144 pairwise comparisons of Euclidean distance in each group. The heavy line at the center of the box is the median value, top and bottom of the boxes define the interquartile range, and the whiskers represent 1.5

times the interquartile range. Euclidean distances between gene expression values of fish in the same population were smaller than distances between fish measured at the same time (p = 2.2e-16, t test).....34

Figure 2-3 Population variation is much greater than the effect of salinity. The fraction of variability explained by population, salinity and interactions between population and salinity for each gene, estimated as the coefficient of determination, r^2. Boxplots represent 34,713 measurements. The heavy line at the center of the box is the median value, top and bottom of the boxes define the interquartile range, and the whiskers represent 1.5 times the interquartile range.....36

Figure 2-4 Osmotic response genes were identified despite the large amount of population variation. Each horizontal bar represents the 1 h or 24 h response of a gene that responded in our experiment (FDR < 0.05) and in at least one other, previously reported case, based on a manual literature search. Response magnitude and standard error (x axis) is based on linear model estimates. Yellow bars represent the responses of fish from Kings Creek, VA (VA), orange bars show the response of fish from Horseshoe Creek, ME (ME2), and red bars show the response of fish from Northeast Creek, ME (ME1).....37

Figure 2-5 Different KEGG pathways were activated at 1 h and 24 h. Bias toward activation or repression was assessed as the fraction of all genes on a path that were induced, less 0.5, the fraction expected to be induced by chance. Red colors (inset) indicate bias toward activation, green colors denote a bias against activation. Bias significance was assessed by binomial tests (*** = p < 0.001). No paths were significant at both 1 h and 24 h.....39

Figure 2-6 Networks of salinity response genes are loosely coupled. Extramural network connectivity, a measure of the connection between a gene set and all other genes, is shown for a random set of 191 genes and the salinity gene set, 191 unique genes that respond significantly to salinity at 1 h or 24 h based on a mixed effect model in Methods. Salinity responsive genes are significantly less connected to other genes than similar sized sets of randomly selected genes are ($p = 0.03$, permutation test).....41

Figure 2-7 Salinity Response Genes Highly Canalized. The relationship between variability (coefficient of variation, y axis) and degree of connection between a gene and all other genes (network cohesion, x axis) for genes in general (black symbols) and salinity responsive genes (blue symbols). Both axes are presented in Z scaled units: zero is the average value for all observations, and differences from zero are expressed in standard units. 83% of salinity-responsive genes were below average (white, dashed line) in coefficient of variation.....43

Figure 2-8 Highly canalized genes cluster by geography. Hierarchical cluster of within-group variability for genes with low variability (coefficient of variation in the first quartile). Height of cluster dendrogram shows Euclidean distance between treatment groups including three exposure durations (0, 1, 24 h) for the Northern clade populations (Northeast Creek, Horseshoe Creek) and Southern clade population (King's Creek). ...44

Figure 3-1 Graphical abstract for Environmental Science and Technology.....51

Figure 3-2 Absolute value of responses in arsenic exposed fish less the absolute response in unexposed fish (Response Delta). Each symbol represents a gene that responded significantly at each time point. The frequency of observing a given Response Delta is shown by kernel density estimations. Red bars reflect the 95% confidence interval of the

mean delta value (black line). A: The impact of arsenic on hypo-osmotic shock. B: The impact of arsenic on hyper-osmotic shock.....58

Figure 3-3 Distribution of correlation (Pearson r) values for four sets of genes significantly differentially expressed in response to (A: hypo-osmotic shock, 1 h; B: hypo-osmotic shock, 24 h; C: hyper-osmotic shock, 1 h; D: hyper-osmotic shock, 24 h). Solid red lines indicate Pearson r value distribution for samples that were not exposed to arsenic. Dashed green lines indicate arsenic exposed samples. 62

Figure 3-4 Net gain or loss in significant correlation for salinity responsive genes in arsenic exposed or arsenic unexposed fish during osmotic shock. A: hypo-osmotic shock, 1 h; B: hypo-osmotic shock, 24 h; C: hyper-osmotic shock, 1 h; D: hyper-osmotic shock, 24 h.....63

Figure 3-5 Transcription factor responses to arsenic (As), 1 h exposure to hypo-osmotic shock, and the combination of arsenic and 1 h exposure to hypo-osmotic shock (1h+As), in $\log_2$ units relative to control. Error bars are 95% confidence intervals. 65

Figure 3-6 Relationship between arsenic impact on number of significant correlations (x axis) and arsenic impact on magnitude of gene expression response (y axis). Each symbol represents a single gene that significantly responded to hypo-osmotic shock at 24 h. Simple linear regression line shown. 69

Figure 3-7 Pathway activation analysis of arsenic, hypo-osmotic shock at 1 h, 24 h, and interaction effects between osmotic shock and arsenic exposure. Cell colors indicate systematic bias toward pathway induction (red) or repression (green) and FDR corrected bias significance based on binomial tests is indicated by asterisks (“*” = 0.05, “**” =

0.01, “***” = 0.001). Interactions between arsenic and osmotic shock appear on the right. 71

Figure 4-1 Graphical abstract for Environmental Science and Toxicology. 73

Figure 4-2 Baseline population expression differences. Gene expression ratios (log_2) between tolerant and sensitive clones in control conditions, plotted as a function of significance ($-log_{10}$ p-value). Genes above the dashed horizontal line achieved an FDR of 0.05. Hexagons colors denote number of observations falling within each area....81

Figure 4-3 Cadmium effects, relative to control, of genes in KEGG pathways that were significantly activated or repressed by cadmium exposure in either the tolerant (y axis) or sensitive (x axis) clone. Grey symbols: pathway genes with FDR > 0.05 in both clones. Red symbols: genes with FDR < 0.05 in both clones. Green symbols: genes with FDR < 0.05 only in the tolerant clone. Blue lines show the linear model fit regression line surrounded by shaded area that represents the 95% confidence interval of the regression. 87

Figure 4-4 Absolute magnitude of cadmium response for all significant genes (FDR < 0.05) in either sensitive or tolerant clones. A: genes with FDR < 0.05 only in one population. B: genes reaching FDR < 0.05 in both clones, *** p = 0.0005 in paired t test. Right: Genes with FDR < 0.05 only in the tolerant clone had significantly larger absolute responses (* p = 0.03, t test) than genes with FDR < 0.05 only in the sensitive clone. ..89

Figure 4-5 Boxplots of absolute difference from control (reaction norm) by toxicant and clone, in log_2 units. Boxes delimit the interquartile range, center lines inside boxes are medians, and whiskers indicate 1.5 times the interquartile range. Green: sensitive clone’s response for genes that responded significantly only in the sensitive clone. Blue:

tolerant clone's response for genes that responded significantly only in the tolerant clone. Pink: sensitive clone's response for genes that were significantly responsive in both clones. Red: tolerant clone's response for genes that were significantly responsive in both clones. 90

Figure 4-6 Co-regulation among cadmium responsive genes. Clone differences in the number of significant correlations, for each gene responding significantly to cadmium. Bar heights represent the number of significant correlations in the tolerant clone less the number of significant correlations in the sensitive clone.....92

Figure 4-7 (Supplemental) Microarray loop design.....93

Figure 5-1 A general model outlining relationships between selective pressures, network cohesion, gene expression reaction norms, gene expression canalization, and effective stress responses. 100

Tables

Table 3-1 1h hypo-osmotic shock genes, arsenic impact on gene-gene correlations (Delta)..67

Table 4-1 Tolerant vs Sensitive at Baseline: Pathway enrichment and activation of three KEGG paths that were significantly enriched (FDR < 0.05) in genes that differed between the two clones in their baseline expression. The number of genes on the path, the enrichment rate, enrichment p value, activation rate, and activation p value is shown. Activation rate specifies the fraction of significant genes that were induced in the tolerant clone relative to the sensitive clone, and the activation p value is based on a binomial test in which the null hypothesis is that the activation rate is 0.5.....82

Table 4-2 Sensitive Response to Cadmium: Pathway enrichment and activation of six KEGG paths that were significantly enriched (FDR < 0.05) in genes that responded to cadmium in the sensitive clone. The number of genes on the path, the enrichment rate, enrichment p value, activation rate, and activation p value is shown. Activation rate specifies the fraction of significant genes that were induced in response to cadmium in the sensitive clone, and the activation p value is based on a binomial test in which the null hypothesis is that the activation rate is 0.5.....83

Table 4-3 Tolerant Response to Cadmium: Pathway enrichment and activation of six KEGG paths that were significantly enriched (FDR < 0.05) in genes responded to cadmium in the tolerant clone. The number of genes on the path, the enrichment rate, enrichment p value, activation rate, and activation p value is shown. Activation rate specify the fraction of significant genes that were induced in response to cadmium the tolerant, and the activation p value is based on a binomial test in which the null hypothesis is that the activation rate is 0.5. 84

Table 4-4 (Supplemental) Estimates of baseline expression differences for genes with annotations matching metallothionein, glutathione, MTF-1, multidrug resistance, cytochrome P450, superoxide dismutase, or heat shock, that differed significantly between sensitive and tolerant clones in baseline expression. *D. pulicaria* gene ID, gene description, $\log_2$ fold change ratio of expression between tolerant and sensitive clones, and p value from linear model comparing baseline gene expression are shown. Strong activation in the tolerant clone ($\log_2$ fold change greater than 0.5) are highlighted in blue, strong repression is highlighted in red 95

Table 4-5 (Supplemental) Estimates of sensitive and tolerant gene expression responses to cadmium relative to control, for all genes with annotations matching metallothionein, glutathione, MTF-1, multidrug resistance, cytochrome P450, superoxide dismutase, or heat shock proteins, *D. pulicaria* gene ID, description, $\log_2$ fold change ratio of expression and p value sensitive clone, followed by $\log_2$ fold change ratio of expression and p value tolerant clone, from linear model estimates. Strong activation in the tolerant clone ($\log_2$ fold change greater than 0.5) are highlighted in blue, strong repression is highlighted in red. 97

1. General Introduction

In 1980, I was injecting male Sprague Dawley rats with nickel and chromium to observe increased strand-breaks and cross-links in DNA. Metal-induced DNA damage was consistent with genotoxicity, and therefore, carcinogenicity (Tsapakos *et al.* 1981; Ciccarelli *et al.* 1981; Tsapakos *et al.* 1983b; a). We used a high dose, 40 mg/kg, causing overt toxic effects including cyanosis (Kim & Na 1990). We speculated privately that carcinogenesis involved mutations in the coding regions of genes associated with cell cycle progression and cell to cell contact, but it was impractical to test this theory. In a way, we were lucky that it was impractical. We wanted cancer to be simple, and we would have been disappointed to learn that the process that leads to cancer is highly diverse. Chromate probably does not cause cancer by recognizing a specific sequence in DNA and causing a mutation leading to unchecked proliferation. Rather, carcinogenesis is a process of evolution involving a very large number of mutations accrued over time (Klein 2013). No single gene explains a large fraction of the cancer phenotype (Vogelstein & Kinzler 2004), and the contemporary focus of cancer research has mostly shifted from attempting to identify how a carcinogen might interact with DNA to understanding how carcinogens affect the pathways governing fundamental cell processes such as cell death and proliferation (Kreeger & Lauffenburger 2010).

In the ensuing years, high throughput gene sequencing, gene expression, and proteomic measurements have quietly changed the philosophical motivation behind research. Formerly, experiments were mostly conducted to falsify a specific hypothesis (Popper 2002). In high dimension gene expression experiments, the goal is less to reject the null hypothesis (that no genes are differentially expressed) and more to identify the genes and hypotheses that can be verified later. For example, in 2005, we observed increased expression of metallothionein

mRNAs in *Daphnia* exposed to cadmium, including novel metallothioneins. These new metallothioneins were individually sequenced, and their expression verified by qPCR (Shaw *et al.* 2007). Still, even at that time, we were also beginning to look beyond individual genes to patterns in the experiment, identifying the fraction of cadmium-responsive genes in various categories to understand the broad outlines of what cadmium does to *Daphnia*.

My first measurements of arsenic's impact on gene expression (Andrew *et al.* 2007; Mattingly *et al.* 2009; Kozul *et al.* 2009b) focused more on systematic effects, and included what we thought of at the time as "new" ways of representing data, such as heatmaps and network diagrams. It is worth remembering that these figures required careful explanation in the early days. In any case, arsenic exposures in mice and zebrafish identified pathway results that were consistent with contemporary epidemiological connections between arsenic in drinking water and susceptibilities such as impaired lung function and respiratory illness (Smith *et al.* 2006; Ghosh *et al.* 2007), and follow up work continues to validate these results both epidemiologically (Farzan *et al.* 2016) and *in vitro* (Goodale *et al.* 2017). In short, integrative and systems biology approaches can identify real effects that stand the test of time.

However, general acceptance that these novel approaches work did not come overnight, in large part because of the multiple hypothesis testing problem. For example, about 1,000 genes out of 20,000 (5%) are predicted to reach ordinary statistical significance (a p value of 0.05) by chance alone. Given this, the question arises of how one distinguishes these predicted false positive genes from the genes that differ in their expression due to experimental conditions. And, if one cannot answer that question easily for a single gene, how can one trust any conclusions drawn from such data, collectively? Lack of confidence was exacerbated by the observation that gene expression profiling performed in different

laboratories did not seem to identify consistent lists of differentially expressed genes (Shi *et al.* 2006). Widespread recognition of this fact led to a sustained debate about whether array measurements of gene expression were scientifically interpretable (Zhang *et al.* 2008). The transition from microarrays to RNAseq was predicted to bring enhanced reproducibility (Wang *et al.* 2009), but gene level concordance remains low between different experiments, partly due to technical difference in the way data are analyzed (Collado-Torres *et al.* 2017) (Consortium 2014). Concordance is enhanced when pathways are considered and data are processed in a consistent way (Gosse *et al.* 2008; Hampton & Stanton 2010; Hampton *et al.* 2012; Oerton & Bender 2017). A second challenge for systems biology and its acceptance is not a consequence of multiple hypothesis testing burden, but the practice of using a large number of statistical approaches, one after another, until the approach that optimizes the number of significant and biologically meaningful results is identified. This approach represents a scientific paradigm shift (Kuhn 1976), differing from the previous practice of strong inference (Platt 1964) and exploratory data analysis (Tukey 1977): it seems to break all the rules. However, it may be that data analysis of this sort is accepted because integration with *a priori* information is an essential feature of the modern paradigm.

Contemporary experimental designs are often less likely today to identify spurious, but statistically significant results, because today's designs are more realistic. As a simple example, my 1980 experiment used a huge dose of injected chromate in male, inbred, litter-mate rats. Today, we would question how well this experiment recapitulates the system we are really trying to study. People are never injected with chromate; they are genetically diverse, male and female; and they are exposed to very low levels of chromate in the air (Junaid *et al.* 2016). The following chapters are in keeping with this trend toward realism, and use genetically diverse organisms from multiple populations responding to environmentally relevant stresses.

In this thesis, I am using the core principles of systems biology (Kitano 2002) to identify the structure, control, and dynamics of gene expression to ask the following question:

> *How do differences in the magnitude, structure, and coordination of gene expression responses to biotic and abiotic stresses observed in natural populations inform our understanding of adaptive responses, such as acclimation to osmotic shock and cadmium tolerance, and toxic responses, such as the interference of arsenic with acclimation to osmotic shock*?

What this Thesis Found

The findings of this thesis can be summarized in two words: Networks matter. More exactly:

> *Coordinated responses to stress with larger magnitudes are associated with cadmium tolerance in Daphnia and more effective acclimation to salinity change in killifish. Tightly controlled genes, organized into networks that share fewer connections with other networks, facilitate osmotic shock responses in killifish. Results gathered in vertebrate and invertebrate wild populations therefore support the views that gene regulatory networks 1) shape the efficacy of biotic and abiotic stress responses, 2) are targeted by toxic effects, and 3) are preserved by selective forces.*

These findings are supported by the results of the three research chapters, summarized below:

Islands of Constraint Amid a Sea of Diversity

Chapter 1, formatted for publication in *Molecular Ecology*, explores individual and population gene expression diversity in the Atlantic killifish, which is known for great diversity at the sequence level (Reid *et al.* 2017). This paper reports that although population

effects explain 42% of the variability in response to changing salinity, specific genes, pathways, and network structures are conserved, even across populations belonging to different clades. This paper verifies a recent report (Shaw *et al.* 2014) that modular networks facilitate acclimation to salinity in the killifish gill. In short, although killifish individuals and populations are highly diverse in their gene expression responses to stress, consistent with the capacity for rapid adaptation that has been observed in this species, diversity is constrained at the pathway and network level.

Arsenic as a Network Toxicant

Chapter 2, formatted for publication in *Environmental Science and Technology*, uses the killifish model from Chapter 1, to explore the effect of arsenic on gene expression responses during salinity acclimation. Arsenic-exposed fish showed muted gene expression responses, and disruption of gene co-expression networks, consistent with the hypothesis that some of arsenic's toxicity may be mediated by its network effects.

Networks Facilitate Cadmium Tolerance

Chapter 3, formatted for publication in *Environmental Science and Technology*, explores population differences in *Daphnia* responding to cadmium and other toxicants, and reveals that gene expression responses to cadmium in a tolerant *Daphnia* population are systematically larger and more coordinated than those observed in a cadmium sensitive population, suggesting that networks facilitate adaptation to cadmium stress.

Methods

Magnitude of Gene Expression Response

One of the systematic differences observed in this thesis is the size of gene expression response. Responses (reaction norms) quantify phenotypic plasticity and are measured as the difference in phenotypic attributes, such as the idealized response to temperature shown in Figure 1.1.

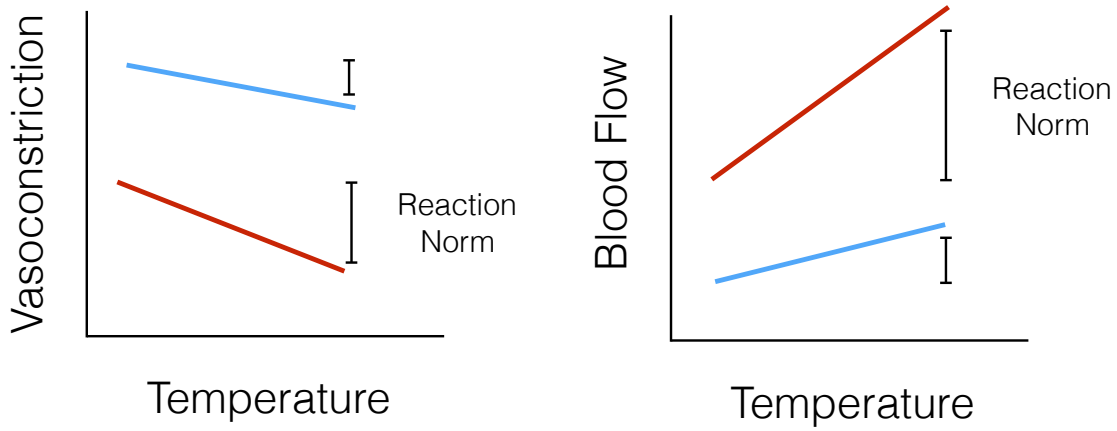

Figure 1-1 Idealized reaction norms for two individuals, "red" and "blue", responding to increased temperature. Red responds more in both changes in vasoconstriction and blood flow, as indicated by the larger reaction norms.

The individual represented by the red line in Figure 1.1 (left) decreases vasoconstriction more with increasing temperature than the individual represented by the blue line. Red also increases blood flow more than blue, consistent with an improved response to heat, and greater tolerance (Charkoudian 2003).

Coordination of Network Response

Systematic differences in the structure of gene regulatory networks were inferred from gene expression data in this thesis, as described below. Figure 1.2 shows simulated gene expression data from which co-regulation might be inferred: each panel represents the expression of a gene measured in 10 different conditions, compared to another gene, measured in the same 10 conditions. For example, the expression of Gene 1 and Gene 2 show little correlation, but the expression of Gene 1 and Gene 3 are correlated across the 10 conditions, as shown by the scatter plot. This co-expression is often represented in a two-dimensional diagram, as a network "connection" between Gene 1 and Gene 3, suggesting that perhaps one gene "regulates" the other. Though correlation does not justify the assertion of causation, it is reasonable to propose that observing significant correlation implies that co-expression is not easily explained by chance. Perhaps Gene 1 and Gene 3 are both targets of the same upstream regulator, for example, and Gene 2 and Gene 4 are targets of a different upstream regulator. Figure 1.2 can be represented as the network in Figure 1.3. The average connectivity in this network is one connection per gene, based on the number of significant correlations. Another way to look at this would be to look at the average absolute correlation between all genes in a network, a measure of general connectivity used by Mackay (Ayroles *et al.* 2009). The subject of how to infer gene networks is an area of active research (Bansal *et al.* 2007), and interconnected genes can be aggregated into subnetworks and larger structures (Yip & Horvath 2007; Marbach *et al.* 2010; Liu *et al.* 2014; He *et al.* 2016).

Two simple metrics were used in the following chapters to quantify structural differences in gene regulatory networks. The first is the average number of significant correlations within a network, intramural connectivity, and the second is the number of connections between networks, extramural connectivity. The latter measure is related to modularity, the extent to

which the behavior of genes in one network is explained by the behavior of genes in another. Networks that are modular (low extramural connectivity) are capable of independent action, and networks of this sort were associated in this thesis with high levels of phenotypic plasticity (Chapter 1). Reduction in intramural connectivity indicates loss of coordination, which was associated with exposure to arsenic (Chapter 2), sensitivity to cadmium (Chapter 3), a phenomenon that has previously been reported as a biomarker of cadmium sensitivity in hexapods (Nota *et al.* 2013).

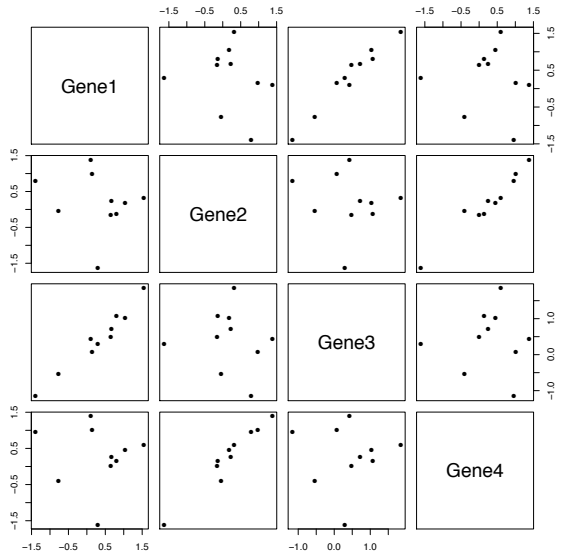

Figure 1-2 Simulated correlations (co-expression) between genes, scatterplot representation Gene 3, Gene 2 correlates with Gene 4, illustrating the concept of gene co-expression.

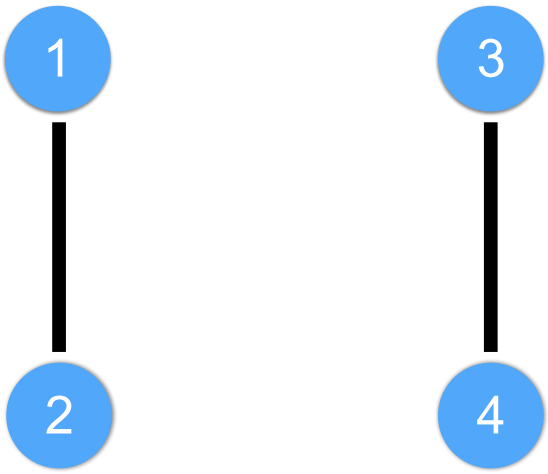

Figure 1-3 Simulated correlations (co-expression) between genes, network representation: Gene 1 and Gene 3, and between Gene 2 and Gene 4, represented as a network diagram.

Pathway Activation

Systematic differences in pathway activation provided an unbiased mechanism to integrate experimentally observed differences in gene expression with prior information in the following chapters.

Unlike inferred network structures, biological pathways rely on a preexisting knowledge base (Khatri *et al.* 2012). Many pathways are also directed graphs, and in this case, causal relationships are typically verified by experiment. In Figure 1.4, arrows indicate that Gene 1 induces the expression of Gene 2, Gene 2 induces Gene 3, etc., as one would expect in a signaling cascade.

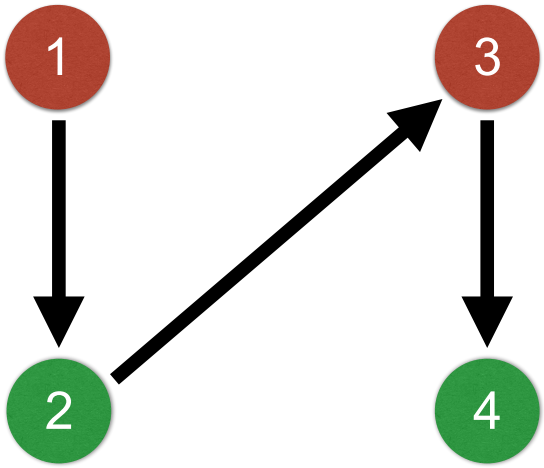

Figure 1-4 Idealized annotated signaling cascade. Gene 1 induces Gene 2, shown by arrows, overlaid with hypothetical gene responses to experimental conditions. Red is induction; green is repression, illustrating a discordance between observed gene expression responses and previously curated gene regulatory relationships.

The simplest bioinformatic analysis of curated paths, one that does not take regulatory relationship into account, is referred to as overrepresentation analysis, and it assesses whether a given path contains more differentially expressed genes than one would predict by chance. For example, if 25% of all genes are differentially expressed in some experiment, one would predict about 1 gene to be differentially expressed in the imaginary path shown in Figure 1.4; if all four were differentially expressed, that would significantly exceed the expected number of genes differentially expressed by chance in this path. But the biological meaning of this significant overrepresentation is less than clear, as shown by the disagreement between the arrows (previously identified regulatory relationships) and the colors (direction of differential expression observed in this hypothetical experiment). If Gene 1 is induced (red) it should induce Gene 2, which should induce Gene 3, etc. However, in this hypothetical experiment, Gene 2 is repressed (green), and Gene 3 is nonetheless induced, yet Gene 4 is repressed.

Therefore, though the path is enriched in differentially expressed genes, the path is both turned on and turned off at the same time, making it complex to include this path in any biological interpretation of treatment effects.

To avoid this problem and identify paths that are systematically induced or repressed, this thesis uses a score of pathway activation based on binomial tests of whether a given path contains more (or fewer) genes that are induced in response to treatment conditions than expected by chance. This test does not rely on significance values of individual genes in its assessment, obviating the need to establish a significance threshold, a parameter to which overrepresentation analysis can be very sensitive.

Tightly Controlled Genes (Gene Expression Canalization)

Finally, systematic differences were observed in the level of variability in different gene sets, as illustrated in Figure 1.5, using simulated data.

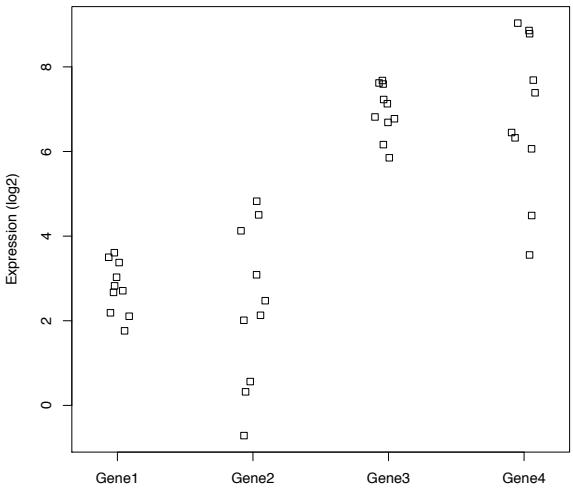

Figure 1-1-5 Simulated gene expression data illustrating genes that are tightly regulated (canalized).

Figure 1.5 shows four simulated gene expression data sets representing the same genes measured in 10 different individuals. Genes 2 and 4 are much more variable (less canalized) than Genes 1 and 3. Variability in gene expression, which is itself variable, is a function of network structure (Manu *et al.* 2009), and relatively tight control (canalization) has been shown to facilitate plasticity in killifish (Shaw *et al.* 2014) and play a role in stress adaptation (Heckel *et al.* 2016). In this thesis, tight control of gene expression is estimated as within-

group coefficient of variation in gene expression, that is, standard deviation divided by the mean.

Background

Killifish and Salinity Acclimation

Osmoregulation is complex in either fresh water or seawater (Figure 1.6) and the two require almost reversed flows of ions. Notably, the fresh water gill actively acquires ions (solid arrows) whereas the salt water gill excretes ions. In fresh water, drinking is limited, but water is passively absorbed through the gill. In seawater, fish drink copiously, absorbing salt and

water, but sodium and chloride are actively excreted through the gill, water is actively reabsorbed from the kidney.

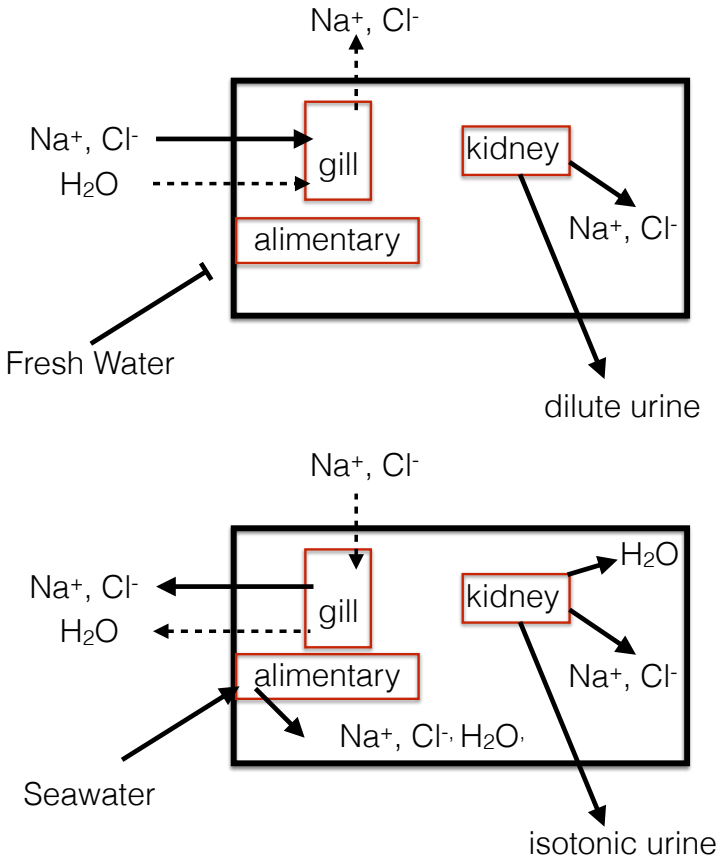

Figure 1-6 Major ion and water flows in fish inhabiting fresh water and seawater environments(Evans 2008).

The Atlantic killifish, *Fundulus heteroclitus,* is well known for its ability to acclimate to changing salinity, a process that involves a complex interplay of organ systems, and the remodeling of the killifish gill over several days (Evans 2010). The gene expression changes that accompany gill transformation not only involve specific genes, such as transcription factors, ion channels, cell-cell contact genes (Whitehead *et al.* 2012b), but also specific network structures, as shown in Figure 1.7. Reduced network complexity and increased gene

canalization was found in the populations that most required phenotypic plasticity, consistent with the hypothesis that natural selection acts on these networks (Shaw *et al.* 2014).

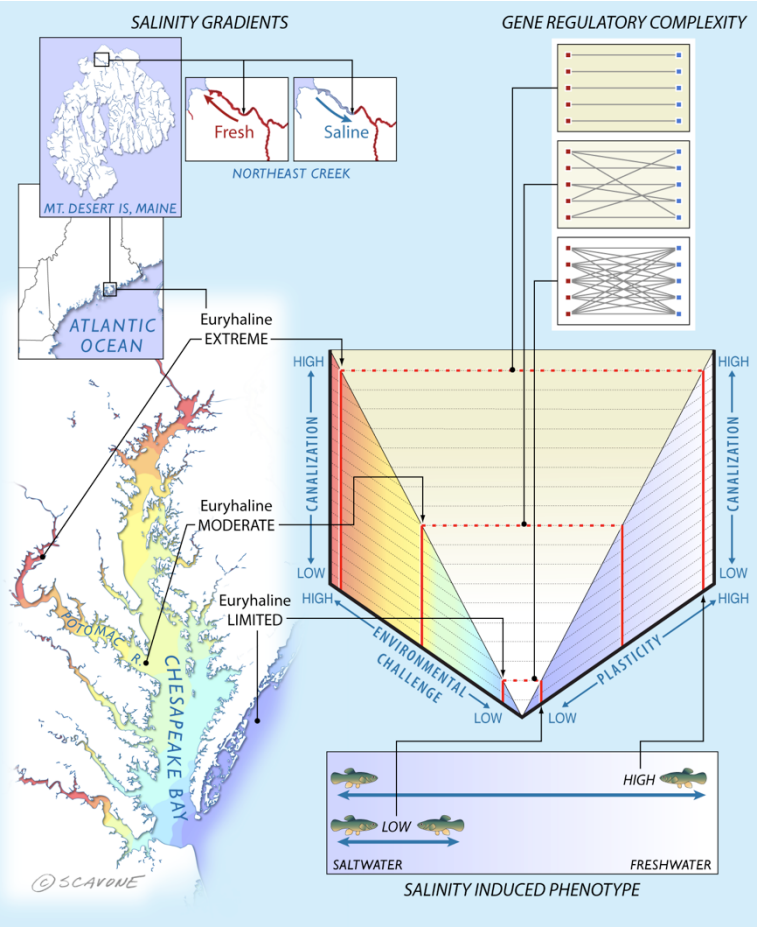

Figure 1-7 A framework relating the ability to acclimate to osmotic shock to gene regulatory network structure. From: Shaw, J. R., Hampton, T. H., King, B. L., Whitehead, A., Galvez, F., Gross, R. H., et al. (2014). Natural selection canalizes expression variation of environmentally induced plasticity-enabling genes. *Molecular Biology and Evolution*, *31*(11), 3002–3015. Reproduced by permission of William Scavone (http://www.kestrelstudio.com/).

Daphnia and Cadmium Adaptation

Daphnids, such as *Daphnia pulex* and *Daphnia magna,* are aquatic crustaceans that have a long history in toxicology and are used as a sentinel species in water quality monitoring (Shaw *et al.* 2008). Individual differences in killifish gene expression, a major feature of Chapter 1, is not easy to address in *Daphnia* for two reasons. First, as a practical matter, *Daphnia* are too small (about 1 mm (Ranta *et al.* 1993)) for standard RNA extraction and profiling on individuals. For this reason, a relatively large number of individual Daphnids (10 – 50) are pooled for gene profiling measurements, eliminating the possibility of measuring individual variability. Second, laboratory stocks of *Daphnia* are derived from a single asexually reproducing individual, as shown on the left of Figure 1.8, resulting in clonal individuals that are not expected to show consistent differences in gene expression.

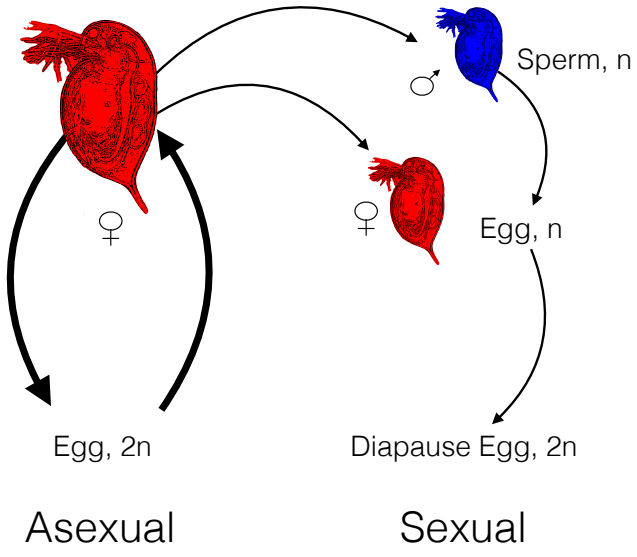

Figure 1-8 Cyclic Parthenogenesis in *Daphnia* : Adapted from: Shaw, J. R., Pfrender, M. E., Eads, B. D., Klaper, R., Callaghan, A., Sibly, R. M., et al. (2008). *Daphnia* as an

emerging model for toxicological genomics. In *Comparative Toxicogenomics* (Vol. 2, pp. 165–328). Elsevier.

Daphnia populations are ecoresponsive (Colbourne *et al.* 2011), rapidly adapting to cyanobacteria (Sarnelle & Wilson 2005), predation (Cousyn *et al.* 2001), temperature (van Doorslaer *et al.* 2009), pesticides (Jansen *et al.* 2011), and metals such a cadmium, tolerance to which can be induced by artificial selection (Ward & Robinson 2005). Chapter 3 looks at gene expression response differences in cadmium-tolerant and cadmium-sensitive populations.

Cadmium Toxicity

Cadmium follows a familiar pattern as an environmental pollutant: human activity is responsible for most of its deposition in the environment, it biomagnifies in food webs (Croteau *et al.* 2005), and its toxicity is most associated with a specific human disease, in this case, kidney disease (Godt *et al.* 2006).

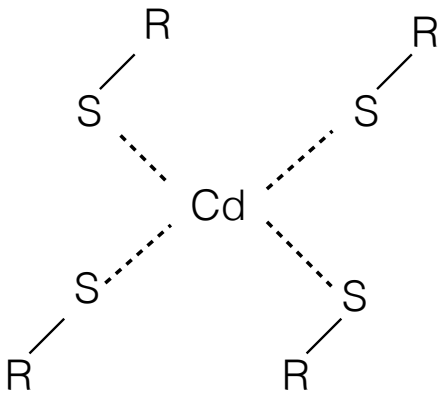

Figure 1-9 Cadmium coordinated by four cysteine residues in an idealized metallothionein(Rubino 2015).

Cadmium in nature is found in mineral ores and coal, e.g., cadmium sulfide, which is released into the environment during mining, smelting, coal burning and the production of cadmium-containing products (Hutchinson & Whitby 1974; Paoliello *et al.* 2002). These products release additional cadmium into the environment as urban waste and sewage sludge. Cadmium is naturally released by the weathering of rocks and volcanic activity (Godt *et al.* 2006).

Once in the environment, cadmium is readily taken up by plants and human exposure occurs mainly through the consumption of high cadmium plants, such as spinach and peanuts, through the consumption of filter feeding shellfish, and by smoking. Over time, cadmium builds up in the body, because it is detoxified through complexation with thiol containing molecules such as metallothionein, as shown in Figure 1.9, but very slowly excreted. The half-life of cadmium in the human kidney, the key organ of concern, is over 10 years (Satarug *et al.* 2017a). Other toxic endpoints include bone malformations, lung cancers, kidney cancers, prostate cancer, neurological issues, hypertension and diabetes (Janssens *et al.* 2009; Benton *et al.* 2011; Satarug *et al.* 2017b), but the current standard, 62 µg/day for a 70-kg person, is based on kidney disease.

The molecular initiating events (Allen *et al.* 2016) that explain the adverse outcomes associated with cadmium remain poorly understood, but have variously been attributed to oxidative stress (Liu *et al.* 2009), interference with DNA repair (Bertin & Averbeck 2006), interaction with free thiol groups (Li 2003), and replacement of essential metals such as zinc, calcium and iron (Martelli *et al.* 2006). Inorganic cadmium can enter cells through calcium, zinc and iron channels (Marchetti 2013), and maybe be released through ABC transporters such as multidrug resistance efflux pumps (Prévéral *et al.* 2009). Detoxification of cadmium by complexation to thiols is highly conserved across species, but varies somewhat by taxon.

Phytochelatins (oligomers of glutathione) play key roles in plants (Sanit di Toppi & Gabbrielli 1999) (Yamaguchi *et al.* 2017), metallothioneins (and sometimes phyochelatins) in animals (Isani & Carpenè 2014), and other thiols such as glutathione participate broadly in many taxa (Delalande *et al.* 2010).

Responses to cadmium have been studied in a variety of natural populations, notably earthworms (Stephen R Stürzenbaum *et al.* 2004; Owen *et al.* 2008; Svendsen *et al.* 2008) such as *L. rubellus*, springtails (Crouau *et al.* 1999; Timmermans *et al.* 2005; Van Straalen & Roelofs 2005; Janssens *et al.* 2009; Roelofs *et al.* 2009; 2010; van Straalen *et al.* 2011; Costa *et al.* 2012; Nota *et al.* 2013) such as *F. candida*, and the water flea (Bodar *et al.* 1990; Stuhlbacher *et al.* 1992; Barata *et al.* 2002; Muyssen & Janssen 2004; Ward & Robinson 2005; Shaw *et al.* 2006; Poynton *et al.* 2006; Shaw *et al.* 2007; Soetaert *et al.* 2007; Connon *et al.* 2008; Haap & Köhler 2009; Clifford & McGeer 2010; De Coninck *et al.* 2013; Asselman *et al.* 2013; Chen *et al.* 2015; Haap *et al.* 2016; Chen *et al.* 2016a).

Arsenic Toxicity

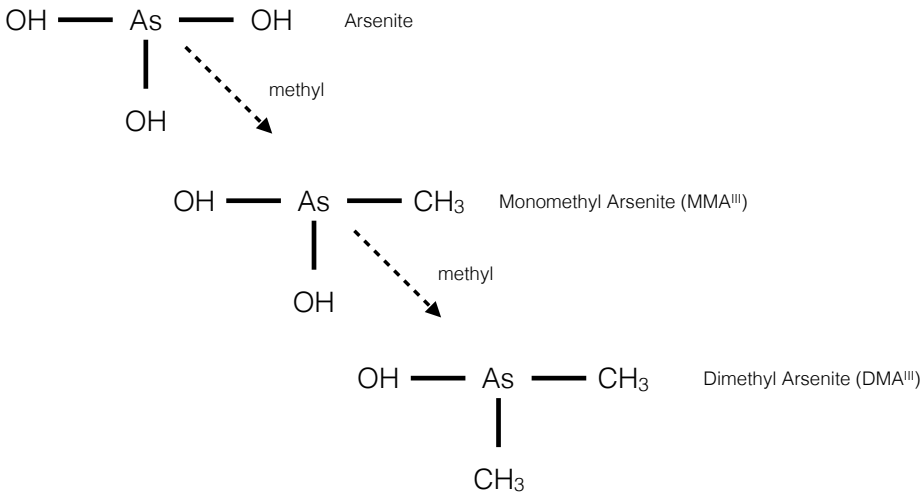

Figure 1-10 Methylation of arsenite (Hughes *et al.* 2011).

Arsenic is well known as a poison used in homicides, but chronic toxic effects, such as its association with the diseases like black foot disease and bladder cancer (Chiang *et al.* 1993), are much less familiar to the lay public. Unlike cadmium, arsenic is naturally present in the environment, is rapidly excreted and not generally biomagnified. Arsenic rapidly enters cells though aquaporin channels (Shinkai *et al.* 2009; Jung *et al.* 2015), and is methylated one or more times before excretion, of the organic species shown in Figure 1.10, a process that may be facilitated by glutathione (Rubino 2015). Trivalent arsenic species, such as arsenious acid, tend to interact with thiols (Hughes 2002), which play a role in arsenic's ability to inactivate up to 200 different enzymes (Ratnaike 2003). Pentavalent arsenic such as arsenate can replace phosphate in biochemical reactions (Tawfik & Viola 2011) leading to a second set of possible initiating events that might lead to toxic endpoints.

Arsenic is concentrated by magmatic and geothermal processes, and geochemistry favorable to arsenic release can be quite variable, leading to natural concentrations that vary enormously even over short distances. Rainwater contains almost no arsenic, lakes and rivers typically contain less than 1 μg/l, and the oceans contain roughly 1 μg/l, well below the current WHO standard of 10 μg/l, whereas highly contaminated groundwater may contain as much as 3,000 μg/l arsenic (Bowell *et al.* 2014).

Human exposure is primarily through the consumption of well water, which may naturally contain 1,000 μg/l arsenic, 100 times the current WHO standard of 10 μg/l. Consuming three liters of such highly contaminated water might cause acute symptoms (Ratnaike 2003). Worldwide epidemiology has shown that lifetime exposure to arsenic at 100 μg/l, the dose used in our killifish experiment in Chapter 2, is associated with an extremely broad range of adverse outcomes that range from intelligence deficits, diabetes, cardiovascular disease, to cancer (Naujokas *et al.* 2013). Infants exposed in utero by mothers consuming water at or

near the 10 μg/l WHO drinking water standard experience adverse immune outcomes in a dose-dependent fashion (Farzan *et al.* 2016). Globally, on the order of 100 million people are chronically exposed to arsenic though drinking water (Argos *et al.* 2012).

2. Conserved Gene Networks Facilitate the Osmotic Shock Response in Killifish

With: Craig Jackson, Stephen P. Glaholt, Dawoon Jung, Andrew Whitehead, Celia Y Chen, Bruce A. Stanton, John K. Colbourne, and Joseph R. Shaw.

Abstract: The Atlantic killifish, *Fundulus heteroclitus,* thrives in estuarine waters where salt concentrations vary enormously over short timeframes and small distances because killifish can acclimate to changes in salinity by altering their gill architecture from a freshwater to a seawater type gill. This phenotypic plasticity is in large part mediated by a coordinated transcriptional response, and natural selection has recently been proposed to conserve gene regulatory networks that facilitate salinity acclimation. In the current study, gene expression responses were measured in three killifish populations to identify differences between individuals and populations at 0 h, 1 h and 24 h, following exposure of saltwater acclimated fish to fresh water. Responses shared among the three populations included induction of c-fos, aquaporin, arginase, claudins, ornithine decarboxylase, and otopetrin, consistent with reports of salinity acclimation in other populations. Osmotic shock systematically induced pathways of genes associated with protein processing in the endoplasmic reticulum at 1 h, and systematically repressed genes associated with RNA transport, spliceosome, and DNA replication at 24 h. Individual differences among fish explained 48% of expression variability, and population effects explained 42% of variability, consistent with high levels of nucleotide diversity observed within and among killifish populations. Salinity-responsive genes formed more independent networks enriched in tightly regulated genes. Northern and Southern populations showed different patterns of canalization. suggesting that regulatory networks are conserved. Collectively, these results suggest that, although the expression of

individual genes is highly variable between killifish individuals and populations, conserved network structures facilitate the osmotic shock response.

Introduction

The study of how genotypic and environmental forces produce phenotypes remains a central challenge of biology. The genomic era has capitalized on the availability of whole genome and whole transcriptome data to reveal the outlines of the molecular basis of evolution, development and disease. Recent work has demonstrated that gene expression differences drive phenotypic difference between individuals (Necsulea & Kaessmann 2014), and that these expression differences are themselves driven by multiple differences in regulatory sequences (Flint & Mackay 2009). Finally, multiple genes, organized in co-expressed modules, ultimately drive phenotype (Ayroles *et al.* 2009).

While it is understandable that much of the pioneering work in these areas involved observing model organisms under stable conditions, certain insights can only be gained by observing outbred individuals under dynamic conditions. For example, reasonable estimates of expression variability within and among populations requires access to outbred individuals in multiple populations. In addition, assessing population differences in gene expression response to realistic environmental stresses requires an experiment performed in a dynamic environment. The Atlantic killifish, *Fundulus heteroclitus,* has long been studied in the context of population differences in environmental stress (Bacanskas *et al.* 2004; Fangue *et al.* 2006; Whitehead *et al.* 2010; 2011b; 2013; Kozak *et al.* 2013; Reid *et al.* 2016), is genetically diverse (Reid *et al.* 2017), and was therefore chosen as the experimental system in which to observe gene expression responses in three populations. Our overarching hypothesis was that acclimation to osmotic shock would reveal significant population differences, consistent with high nucleotide diversity and rapid adaptation to environmental stress.

Killifish thrive in estuarine environments along the Atlantic coast of North America from northern Florida to the Saint Lawrence River. Some populations of killifish are adapted to withstand a wide range of dissolved oxygen concentrations, temperatures and salinities. The ability to tolerate changing salinity and other environmental stressors likely plays a role in making killifish the most abundant vertebrate species in Atlantic coast estuarine environments (Whitehead *et al.* 2012b). Not surprisingly, killifish maintain large effective population sizes and harbor some of the largest known levels of genomic variability (Reid *et al.* 2017), which is predicted to increase variability at the gene expression level. Oleksiak reports that 18% of killifish genes are expressed at significantly different levels in different individuals from the same population (Oleksiak *et al.* 2002; 2005). Although most expression differences are probably neutral (Whitehead & Crawford 2006), gene expression variability plays a key role in adaptation (Enard *et al.* 2002; Khaitovich *et al.* 2004; 2005; 2006).

Killifish populations broadly belong to Northern or Southern clades that separate at the Hudson River in New York, US. (Schulte 2007; Duvernell *et al.* 2008; Whitehead 2009). Northern killifish, and populations from the Southern clade that have adapted to live in freshwater niches, exhibit the greatest ability to acclimate to changes in salinity (Whitehead *et al.* 2012a; Whitehead 2012; Whitehead *et al.* 2012b; 2013; Kozak *et al.* 2013; Shaw *et al.* 2014; Brennan *et al.* 2015). Natural selection appears to shape the gene regulatory networks that facilitate acclimation, yielding networks that are tightly controlled, yet loosely coupled to upstream regulators (Shaw *et al.* 2014).

The underlying biology of killifish salinity acclimation provides the context in which shared and divergent population differences can be understood. Acclimation is most evident in the gill, the major osmoregulatory organ in euryhaline teleosts, which remodels following exposure to hyper-tonic or hypo-tonic stress (Hwang & Lee 2007; Whitehead *et al.* 2011a;

2012b; Kang *et al.* 2013; Kozak *et al.* 2013; Kültz *et al.* 2013; Brennan *et al.* 2015; Cozzi *et al.* 2015; Kültz 2015). In saltwater conditions, mitochondrion rich cells (MRC) in the gill pump sodium and chloride out of the bloodstream and shed excess chloride into the external environment. In reduced salinity, killifish rearrange gill morphology to prevent water accumulation and dramatically reduce sodium and chloride secretion (Evans 2010). Although remodeling of gill tissue takes days (Katoh 2003), chloride secretion is radically reduced upon transfer to freshwater within 10 minutes (Guan *et al.* 2016). Other molecular events occur within hours, including changes in polyamine metabolisms and c-fos signaling (Brennan *et al.* 2015; Guan *et al.* 2016). One predicts that gene expression responses in our three populations will reflect what is already known about salinity acclimation in the gill, and that gene expression responses will be largely concordant, especially in killifish belonging to the same clade.

Killifish gill remodeling is a canonical example of phenotypic plasticity: many genes are differentially expressed (Whitehead *et al.* 2012b; 2013; Kozak *et al.* 2013) sometimes under the influence of regulators such as ERBB2, TNF and ERK (Whitehead *et al.* 2012b), but the role played by gene regulatory networks during phenotypic plasticity is not well understood. However, recent work in our lab has shown that plasticity-enabling genes identified in a single Northern-clade killifish population form streamlined, modular networks composed of tightly regulated (highly canalized) genes (Shaw *et al.* 2014). This suggests that specific network structures might facilitate phenotypic plasticity. To test this hypothesis, I therefore assessed network structure of plasticity-enabling genes in three killifish populations.

Materials and Methods

Unless otherwise noted, statistics were performed in R (Team 2014a). PCA plots were created using the ecodist package (Goslee & Urban 2007). Venn diagrams were created using gplots (Gregory *et al.* 2016).

Killifish Populations and Exposures

All work involving animals conformed to Institutional Animal Care and Use guidelines approved by MDI Biological Laboratory. *F. heteroclitus* populations can be classified based on microsatellite DNA as belonging to either a Northern or Southern clades along the Atlantic coast (Whitehead 2009). Northern clade fish were collected by minnow trap from two locations less than 40 km apart: Northeast Creek (Bar Harbor, Maine) and Horsehoe Creek (Brooksville, Maine). Both sample locations are brackish at high tide, and experience freshwater conditions at low tide. Southern clade *F. heteroclitus* were obtained from Kings Creek, Virginia, courtesy of Richard DiGulio of Duke University, from an area that is brackish at high tide.

Fish were maintained outdoors under natural light cycles (photoperiod 15:9-h light:dark) and fed commercial flake food (48% protein, 9% fat; Tetracichlid, Tetra, Blacksburg, VA) once a day. Aquaria were aerated and no appreciable variation in general water quality parameters such as ammonia, pH, salinity, and temperature was detected. All fish were held in aquaria containing running seawater (pH 8.1 ± 0.4; salinity 33 ± 0.5‰, 15 °C) for at least 2 weeks to ensure acclimation (Stanton *et al.* 2006). Fish undergoing hypo-osmotic challenge were abruptly transferred from seawater [30 ppt] to freshwater [1 ppt], and control fish (0 h in freshwater) were maintained in seawater for the duration of the experiment. Fish were sampled at 0, 1, and 24 h after transfer to fresh water. Fish were anesthetized in tricane (0.2

g/l), euthanized by double pithing (Shaw et al. 2010), and gill tissue was immediately isolated, rinsed, and placed in RNAlater (Ambion) according to manufacturer's recommendations.

A total of 36 fish were used: three times points for each of the three populations (nine experimental groups) with four fish in each experimental group.

RNA Isolation

Total RNA was isolated from gill using RNAeasy kits (Qiagen, Valencia, CA) as described in Shaw et al. (2014). Tissue samples were homogenized on ice in lysis buffer containing guanidine isothyocyanate using a Tissue-Tearor (Biospec Products, Barlesville, OK), and RNA was precipitated using an equal volume of ethanol. The RNA was loaded on glass fiber columns, washed, and eluted with nuclease free water. Samples were then DNAase treated (DNAfree, Ambion), quantified using spectrophotometric optical density (OD260/280) measurements (Nanodrop ND-1000, ThermoFisher Scientific, Waltham, MA) and their integrity was assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies, Wilmington, DE). All samples used in genome expression studies achieved a RNA integrity number (RIN) score >7.

RNA Quantitation by Microarray

RNA was amplified using MessageAmp II kits (Ambion) and aRNA labeled using a single chanel design with Cy3 dye using NimbleGen Labeling Kit (Roche NimbleGen). Labelled samples were randomly hybridized using the Hybridization Kit (Roche NimbleGen) to a custom NimbleGen array whose design and annotation has been previously described (Shaw *et al.* 2014). Briefly, this platform interrogates 135,000 probes that map to 34,713 gene

models and 16,104 unique *F. heteroclitus* genes from a high-quality killifish reference genome (Reid et al., 2017). Hereafter, the 34,713 gene models are referred to as genes. Following hybridization, raw data were extracted using NimbleScan v2.4 software (Roche NimbleGen) and quantile normalized across arrays.

Gene Expression Data Analysis

When more than one probe measured the expression of a gene, the probe having the highest median expression was used to represent the expression of that gene. Full linear models, which included factors for population (Northeast Creek, Horseshoe Creek, or King's Creek), time in freshwater (0 h, 1 h, or 24 h), and interactions between population and time in freshwater, were computed for each gene. The explanatory power of factors was calculated by subtracting the r squared of a model lacking that factor from the full model. Independent linear models were run for each of the three population as a function of time in salinity.

Gene Set Activation Analysis

The quantile normalization used in this analysis guarantees that all samples share identical medians. Therefore, a randomly selected gene has a 0.5 probability of being relatively induced in any comparison between two samples. Under the null hypothesis, genes selected by virtue of their membership in known biological processes, e.g., genes belonging to KEGG (Kanehisa & Goto 2000; Kanehisa *et al.* 2007) will also show a rate of relative induction of about 50%. I used binomial tests to assess significant deviations from the null expectation of 50% for each KEGG path, and corrected significance using the false discover rate (FDR) of Benjamini and Hochberg (Benjamini & Hochberg 1995). Activation was expressed as the difference between the observed rate of relative gene induction and 0.5.

Extramural Network Cohesion

Ayroles, Mackay et al (Ayroles *et al.* 2009) used average, pairwise, absolute Pearson correlation as a measure of average connectivity between genes. In this study, I used a similar metric, average pairwise r^2. For each gene belonging to a specific set, correlations were sequentially calculated between expression of that gene and the expression all other genes not belonging to the given set. These pairwise Pearson correlations were squared to yield r^2, the coefficient of determination, and averaged for each gene, to yield extramural network cohesion. Intuitively, extramural cohesion is inversely proportional to the inferred network modularity of a gene set, because it estimates the average strength of upstream and downstream the regulatory relationships that exist between a gene set and all other genes. The significance of differences in extramural network cohesion were assessed by resampling. Specifically, pairs of random samples of N genes were drawn 1,000 times, and mean r^2 calculated between each gene in each random set, and all other genes not in that set, yielding extramural network cohesion for 1,000 pairs of random gene sets of size N. For each pair of random samples, differences of median r^2 were calculated, yielding a null distribution representing the differences in extramural network cohesion likely to be observed by chance, allowing p values to be associated with observed differences in median extramural network cohesion.

Gene Canalization

Treatment group coefficient of variation (COV) was calculated for each gene in each of the nine groups (i.e. three populations, three time-points) and averaged across all treatment groups, as a measure of gene canalization. The relative canalization of each gene was calculated as a z-score, that is, the COV of each gene, measured in standard units compared to the overall mean COV. This perspective allowed genes to be dichotomized as “high” or

"low" COV based on z-score sign. Significant biases in z-score sign were identified with binomial tests.

Results

High Gene Expression Variability Within and Among Killifish Populations

Principal components analysis of sample data reveals three distinct groups in Figure 2.1, one for each population. This is equivalent to saying that fish from the same population regulate the expression of their genes more similarly than fish from different populations. Interestingly, fish exposed to fresh water for the same number of hours (digits, in Figure 2.1) do not cluster together, even within the same population (color).

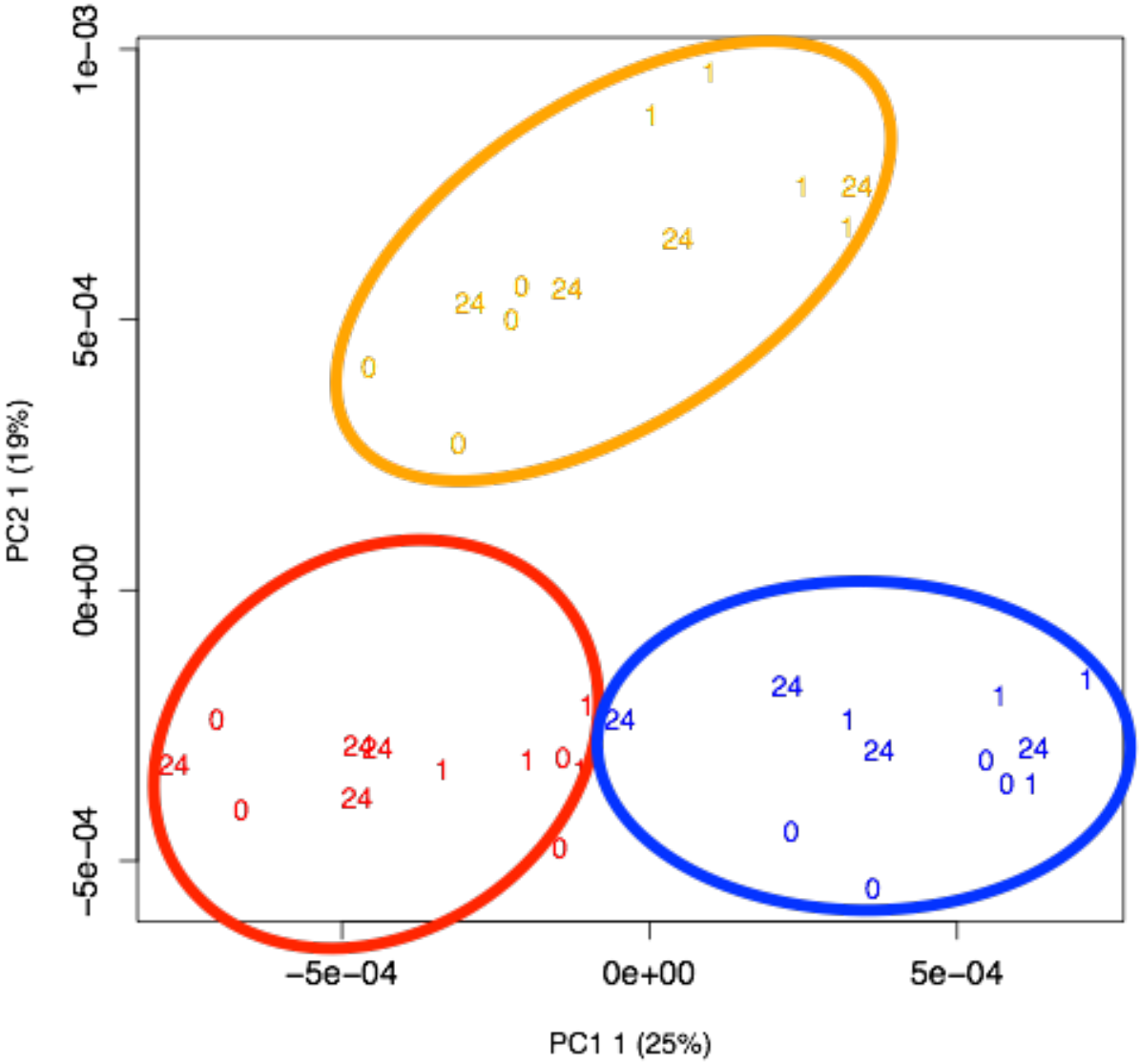

Figure 2-1 Gene expression variation distinguishes populations. Principal components analysis of log transformed, normalized gene expression values for 36 saltwater-acclimated killifish from King's Creek, VA (yellow), Northeast Creek, ME (blue) and Horseshoe Creek, ME (red) following exposure to fresh water for 0, 1, or 24 hours. X axis shows loadings associated with the first principal component (PC1) which accounts for 25% of the variability of gene expression; the second component (y axis) accounts for 19% of variability.

This indicates very high levels of individual gene expression variability within populations. Since the first two principal components account for less than half of the total variability, I used pairwise Euclidean distance, which uses a sum of squares measure across all genes, to compare gene expression profiles across all pairwise comparison between fish. As shown in Figure 2.2, statistical analysis of pairwise comparisons verified that fish from the same population were more similar to each other than fish exposed to fresh water for the same number of hours (t test p value < 2.2e-16) suggesting that population effects are a more important factor determining gene expression than the treatment conditions that were used.

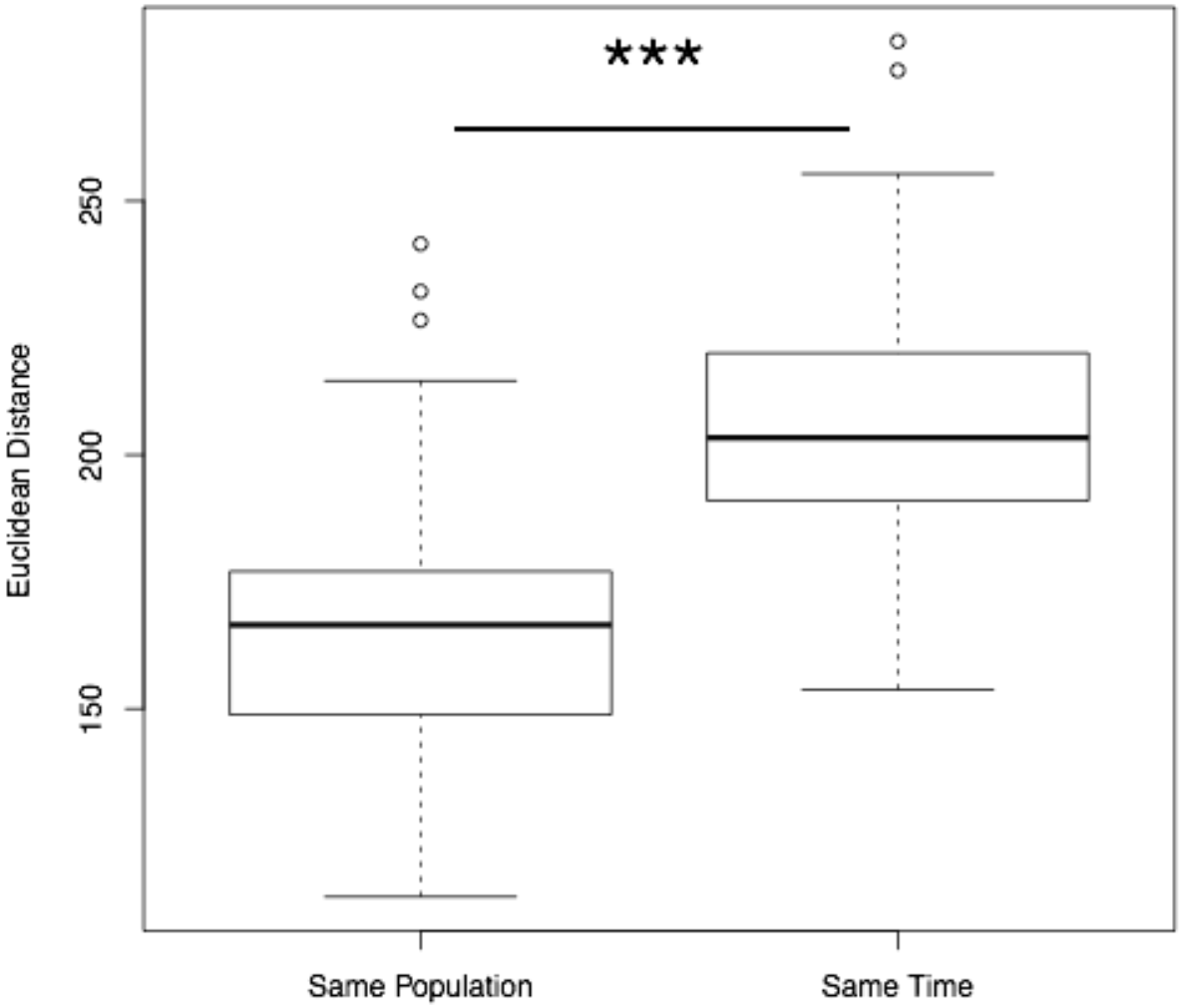

Figure 2-2 Fish from the same population are significantly more alike. Pairwise Euclidean distance measures for fish from the same population, but different hours of exposure to fresh water (left) or fish exposed to fresh water conditions for the same time, but not from the same population. Boxplots represent 144 pairwise comparisons of Euclidean distance in each group. The heavy line at the center of the box is the median value, top and bottom of the boxes define the interquartile range, and the whiskers represent 1.5 times the interquartile range. Euclidean distances between gene expression values of fish in the same population were smaller than distances between fish measured at the same time ($p = 2.2e-16$, t test).

To quantify the amount of variability accounted for by individual differences, populations, and exposure duration, I used r^2 values derived from the mixed effect linear models described in Methods. Briefly, dropping individual terms from the full model reduces the overall r^2 in proportion to the explanatory contribution of that term. In this analysis, I saw that population explains much more variability (33%) than salinity (9%) or interactions between salinity and

population (10%) as shown in Figure 2.1.

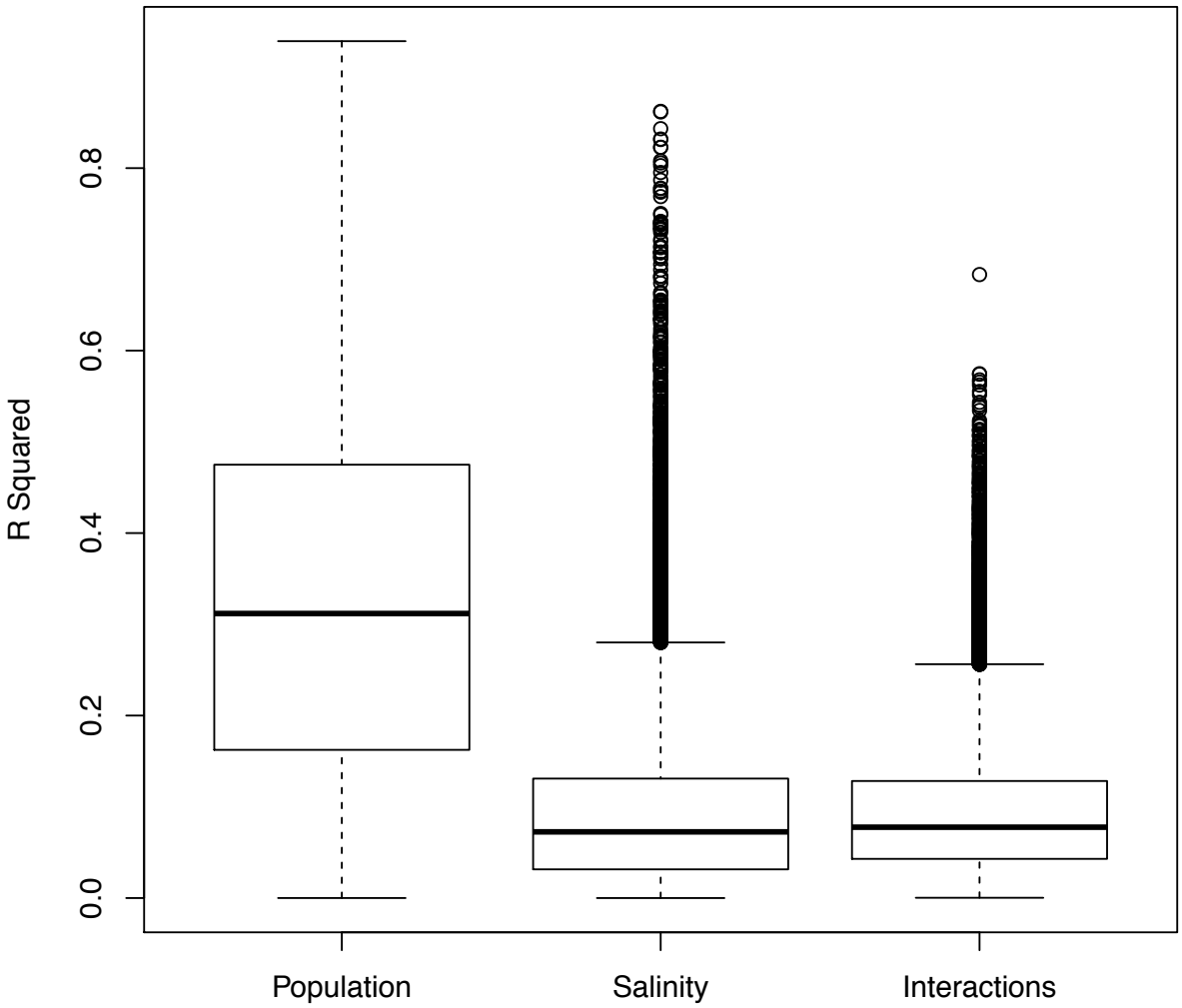

Figure 2-3 Population variation is much greater than the effect of salinity. The fraction of variability explained by population, salinity and interactions between population and salinity for each gene, estimated as the coefficient of determination, r^2. Boxplots represent 34,713 measurements. The heavy line at the center of the box is the median value, top and bottom of the boxes define the interquartile range, and the whiskers represent 1.5 times the interquartile range.

Almost half of the variability (48%) was unexplained by model factors, consistent with high levels of inter-individual variation. Interestingly, salinity and interactions between salinity and population each explained about 9% of the variability in the data, meaning that the population-specific responses to osmotic shock were similar in magnitude to responses shared between populations.

Known Osmoregulation Responses were Shared by Populations

Although most responses to reduced salinity were population-specific, many genes with previously identified roles in teleost salinity acclimation showed very similar responses in the three populations (Figure 2.4).

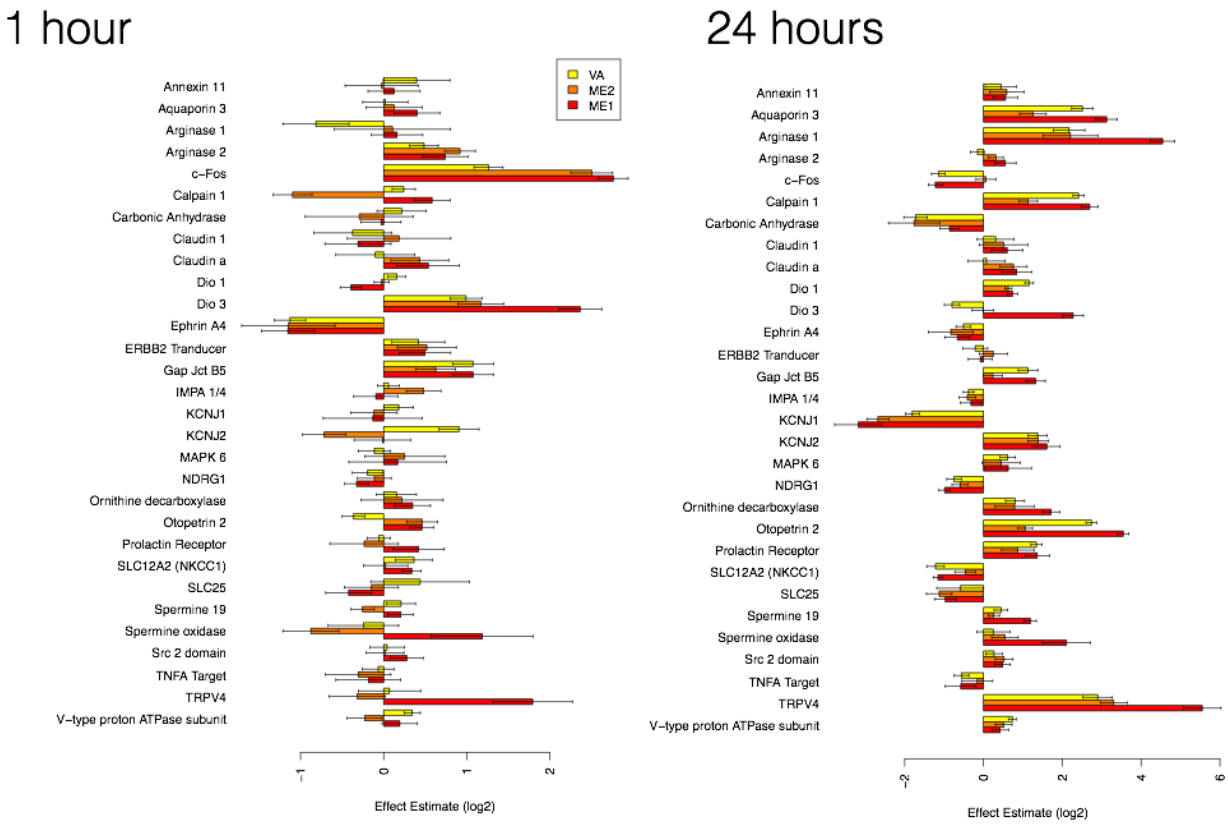

Figure 2-4 Osmotic response genes were identified despite the large amount of population variation. Each horizontal bar represents the 1 h or 24 h response of a gene that responded in our experiment (FDR < 0.05) and in at least one other, previously

reported case, based on a manual literature search. Response magnitude and standard error (x axis) is based on linear model estimates. Yellow bars represent the responses of fish from Kings Creek, VA (VA), orange bars show the response of fish from Horseshoe Creek, ME (ME2), and red bars show the response of fish from Northeast Creek, ME (ME1).

Figure 2.4 was constructed using a mixed effect model to identify salinity responsive genes, searching the literature for previous reports of these genes in the context of salinity acclimation, and plotting the independent response of each population, using effect estimates and standard errors from independent linear models, as defined in Methods. Not surprisingly, many genes show highly consistent behavior, since the genes were selected on the basis of consistent population behavior by the mixed effect model, but what is more interesting, is the extent to which genes show characteristic early and late phase responses across a broad spectrum of functions including genes like TRPV4 that sense osmotic change (Fiol & Kültz 2007), genes like deiodinase (Whitehead *et al.* 2011a), prolactin and c-fos (Kültz 2012) that signal this change to other genes, and genes like arginase, aquaporin and ornithine decarboxylase that respond (Guan *et al.* 2016) to osmotic shock.

Different KEGG Pathways were Activated at 1 h and 24 h

The common response shared by the three populations extends beyond the genes shown in Figure 2.4 and includes systematic regulation of the functional pathways shown in Figure 2.5.

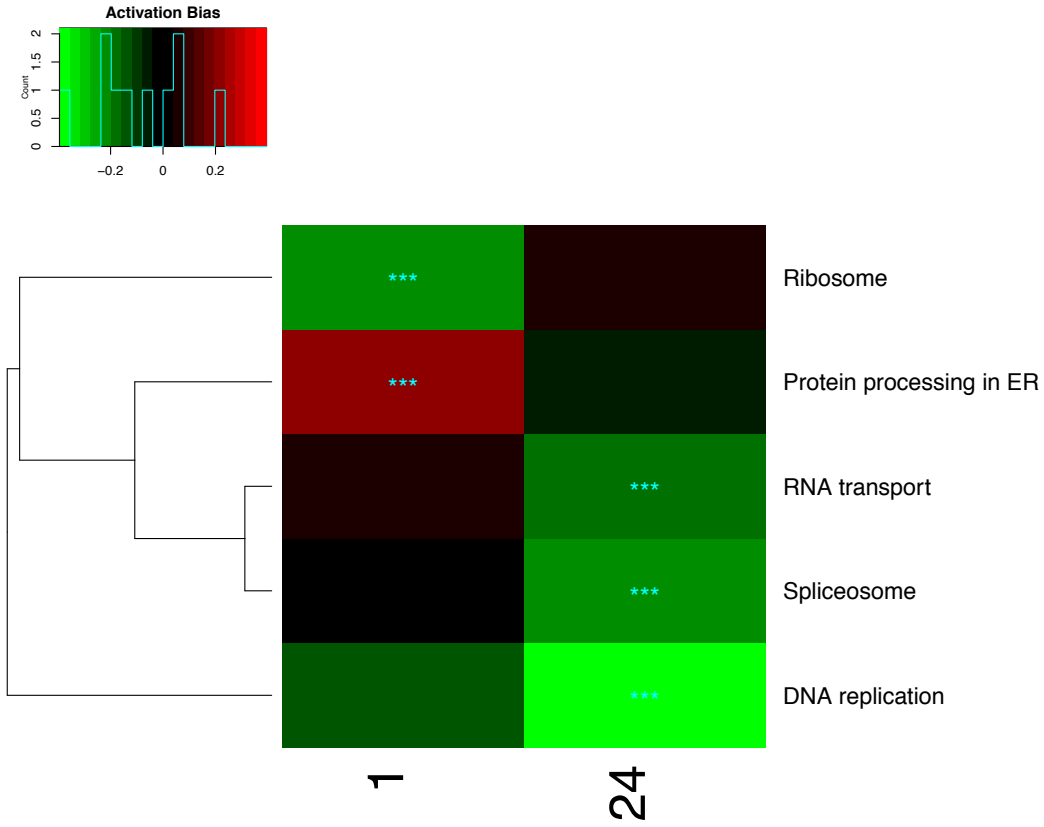

Figure 2-5 Different KEGG pathways were activated at 1 h and 24 h. Bias toward activation or repression was assessed as the fraction of all genes on a path that were induced, less 0.5, the fraction expected to be induced by chance. Red colors (inset) indicate bias toward activation, green colors denote a bias against activation. Bias significance was assessed by binomial tests (*** = $p < 0.001$). No paths were significant at both 1 h and 24 h.

Five KEGG pathways (Kanehisa & Goto 2000; Kanehisa *et al.* 2007) were either activated (red) or repressed (green) based on gene set activation analysis as defined in Methods. Briefly, effect estimates at 1 hour or 24 hours are equally likely to be positive or negative relative to control under the null hypothesis, but many paths were systematically biased toward either activation or repression. For example, the KEGG path "Protein processing in the endoplasmic reticulum" (ER) was biased toward induction at 1 h. Of the 84 killifish genes that mapped to zebrafish genes on this path, 60 (71%) were nominally induced based on effect estimate in the linear model, 21% more than the 50-50 split expected by chance. The

bias toward activation is therefore +0.21, and achieves an unadjusted p value of 0.0001 in a binomial test. Similar calculations were made for all 160 KEGG paths available for zebrafish, at both 1 hour and 24 hours. Ribosome, RNA transport, spliceosome, and DNA replication were significantly biased toward repression, $p < 0.001$ in all cases. Importantly, paths that were induced or repressed at either 1 h or 24 h showed different behavior at the other time point.

Networks of Salinity Response Genes were Loosely Coupled

We used a simple measure of network structure, extramural network cohesion as defined in Methods, to test whether salinity response genes were significantly less connected to other genes than one would predict by chance, since loose coupling (increased modularity) has been previously observed in genes that facilitate phenotypic plasticity (Shaw *et al.* 2014).

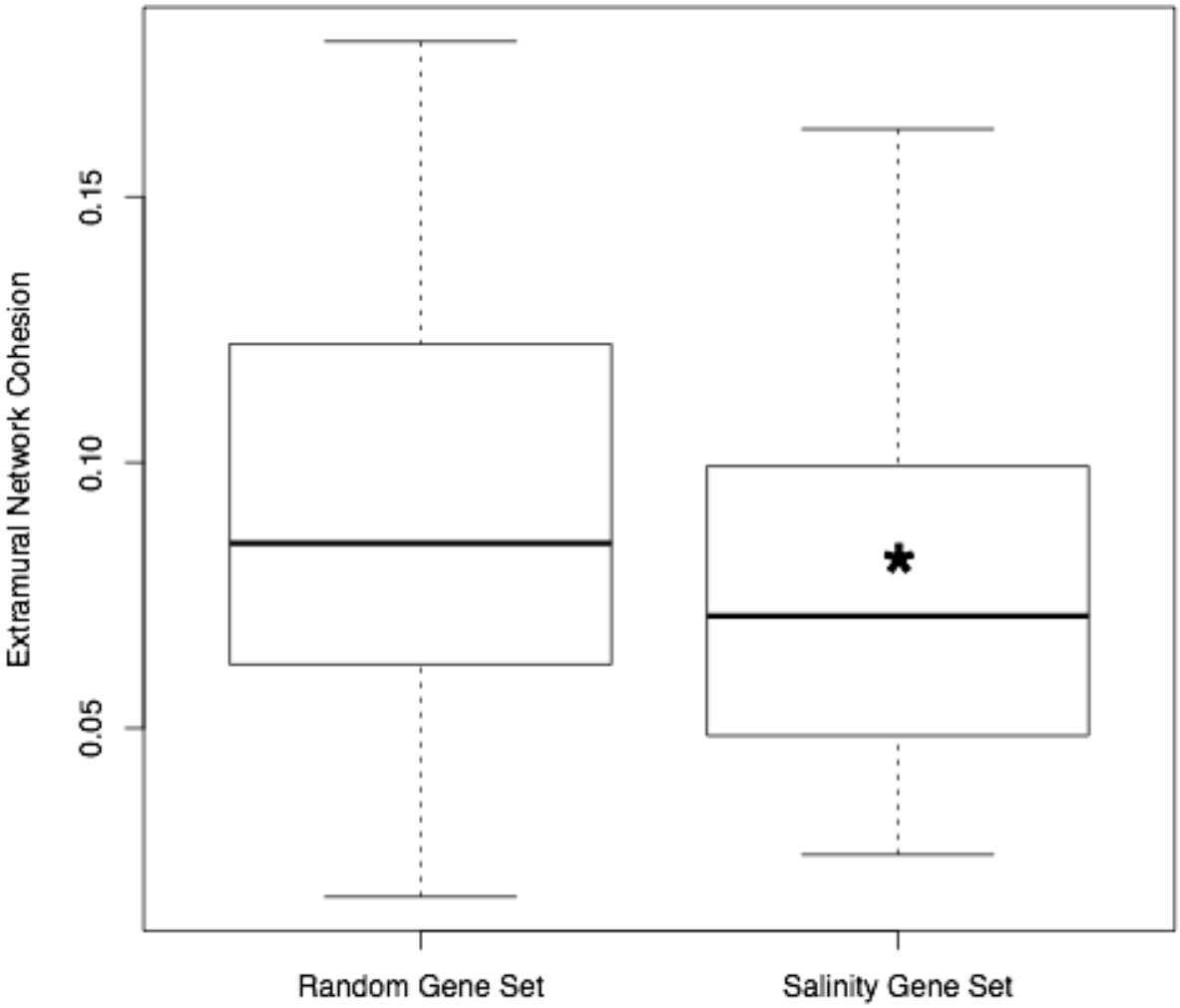

Figure 2-6 Networks of salinity response genes are loosely coupled. Extramural network connectivity, a measure of the connection between a gene set and all other genes, is shown for a random set of 191 genes and the salinity gene set, 191 unique genes that respond significantly to salinity at 1 h or 24 h based on a mixed effect model in Methods. Salinity responsive genes are significantly less connected to other genes than similar sized sets of randomly selected genes are ($p = 0.03$, permutation test).

Figure 2.6 compares the extramural network cohesion of a random set of 191 genes with the extramural network cohesion of the 191 unique genes that responded significantly to hypo-

osmotic shock at 1 h or 24 h. In both cases, average extramural network cohesion was less than 0.1, even for randomly selected genes, indicating that the behavior of a randomly selected gene accounts for less than 10% of other genes. Salinity responsive genes showed less extramural network cohesion than random sets of genes ($p = 0.03$, permutation test).

Salinity Response Genes were Highly Canalized

Previous observations of gene expression in Northeast Creek fish acclimated to fresh water and exposed to salt water showed that plasticity enabling genes were not only more loosely coupled to the behavior of other genes, but also that plasticity genes were more tightly regulated, as evidenced by lower average within-group coefficient of variation (Shaw *et al.* 2014). Based on this observation, I predicted that salinity response genes would also have less variability than genes in general, and this is correct, as shown in Figure 2.7.

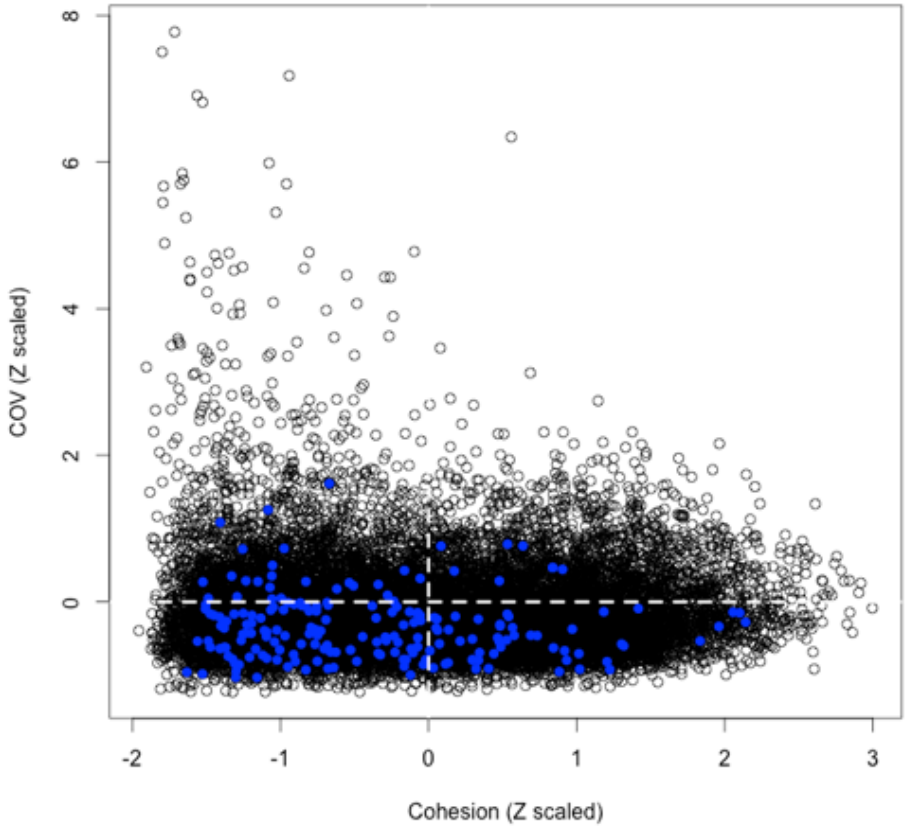

Figure 2-7 Salinity Response Genes Highly Canalized. The relationship between variability (coefficient of variation, y axis) and degree of connection between a gene and all other genes (network cohesion, x axis) for genes in general (black symbols) and salinity responsive genes (blue symbols). Both axes are presented in Z scaled units: zero is the average value for all observations, and differences from zero are expressed in standard units. 83% of salinity-responsive genes were below average (white, dashed line) in coefficient of variation.

Indeed, 83% of salinity genes had a coefficient of variation less than the median, as shown in Figure 2.7, far more than would be expected by chance ($p < 2.2$ e-16, binomial test).

Highly Canalized Genes Clustered by Geography

Highly canalized genes, e.g. those in the first quartile of average coefficient of variation across all treatment groups, had similar levels of canalization in geographically proximate populations, across fish in different treatment groups (Figure 2.8).

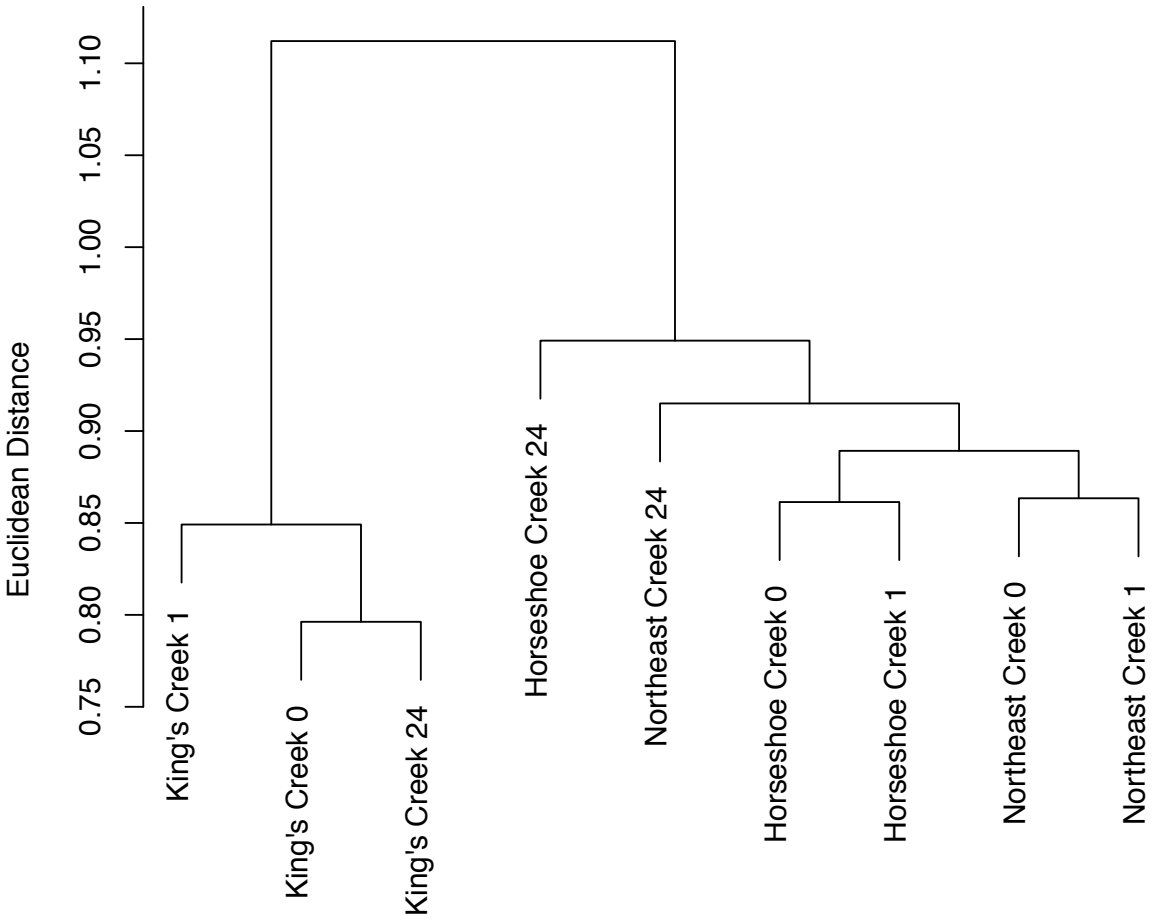

Figure 2-8 Highly canalized genes cluster by geography. Hierarchical cluster of within-group variability for genes with low variability (coefficient of variation in the first

quartile). Height of cluster dendrogram shows Euclidean distance between treatment groups including three exposure durations (0, 1, 24 h) for the Northern clade populations (Northeast Creek, Horseshoe Creek) and Southern clade population (King's Creek).

We observed similarity in COV among fish for genes that were most tightly controlled. Fish from Northeast Creek and Horseshoe creek, Maine, exhibited more similar levels of canalization compared to fish from the King's Creek, Virginia.

Discussion

Here I found that three killifish populations are both convergent and divergent in their responses to osmotic shock. Large population differences in gene expression have previously been observed in mouse populations separated for as little as 3,000 years (Bryk *et al.* 2013), closely related flycatcher populations (Uebbing *et al.* 2016), and other studies of killifish (Oleksiak *et al.* 2002; Crawford & Oleksiak 2007). The high variability we observed in gene expression is consistent with high nucleotide diversity (Reid *et al.* 2017), and predicted to facilitate the rapid adaptation observed in killifish (Reid *et al.* 2016). At the same time, killifish gene expression variability is limited by specific genes, pathways, and regulatory networks that appear to facilitate phenotypic plasticity during hypo-osmotic shock, consistent with the finding that natural selection canalizes expression variation of environmentally induced plasticity-enabling genes during hyper-osmotic shock in killifish (Shaw *et al.* 2014).

Amid Population Variability, Familiar Responses are Conserved

Approximately 200 genes in this experiment responded similarly to hypo-osmotic shock across our three *F. heteroclitus* populations. Many of these (Figure 2.4) were genes with a previously identified role in teleost salinity acclimation (Kultz & Avila 2001; Marshall *et al.*

2005; Fiol *et al.* 2006a; b; Fiol & Kültz 2007; Whitehead *et al.* 2011b; 2012b; Kültz 2012; Kültz *et al.* 2013; Lam *et al.* 2014; Brennan *et al.* 2015; Kültz 2015; Velotta *et al.* 2017). Gene expression profiles included genes associated with all aspects of the required osmotic shock response, including sensing it, surviving it, and altering gill parameters such as ion and water flow to survive salinity change.

Robust, transient induction of the immediate early response gene, c-fos, occurred in both Maine and Virginia populations of *F. heteroclitus* at 1 hour (Figure 2.4, left) a response shared with *Fundulus grandis* exposed to hypo-osmotic shock (Guan *et al.* 2016). Arginase 2, deiodinase 3, ERBB2 transducer, and gap junction B5 were also induced at 1 hour, consistent with both sensing of osmotic change, and responding to it. Arginase is involved in polyamine synthesis, low molecular weight cations that limit the impact of osmotic shock on cell volume (Guan *et al.* 2016). Deiodinases activate thyroid hormone and have been previously reported to be differentially expressed during killifish salinity acclimation (Whitehead *et al.* 2011b). ERBB2 is a receptor tyrosine kinase that appears to function as an important upstream regulator in killifish salinity response (Whitehead *et al.* 2011b), and gap junction proteins are known to be induced in teleost fish responding to hypo-osmotic shock (Lam *et al.* 2014). Alterations to gap junctions and other cell adhesion proteins likely influence pericellular water flow and tissue remodeling in the gill. Genes repressed at 1 h included Ephrin A4, a component of axonal guidance signaling that is regulated during gill tissue during osmotic stress (Lai *et al.* 2015).

Significant responses shared between populations were more numerous at 24 hours and included the induction of annexin, aquaporin, arginase, calpain, claudins, deiodinase, a potassium channel, mapk6, ornithine decarboxylase, otopetrin, prolactin receptor, spermines, src 2 domain, TRPV4, and a v-type proton assembly unit. Induced genes at 24 hours

therefore include examples of genes with known roles in sensing, crisis survival and long-term acclimation to osmotic shock. Induced sensing components notably include TRPV4, a stretch activated calcium channel (Seale *et al.* 2012), that mediates prolactin release, which is thought to signal the need for gill remodeling in hypo-osmotic conditions (Kültz 2015). Calcium responsive genes such as otopetrin, annexin, and calpain were also induced, and they may counterbalance cell swelling induced by hypo-osmotic shock (Brennan *et al.* 2015). Increased spermine synthesis, evidenced by regulation of ornithine decarboxylase and spermines, serves a protective role during hypo-osmotic shock (Guan *et al.* 2016). Inositol also plays a protective role during osmotic shock (Kozak *et al.* 2013), and inositol monophosphatases are regulated in killifish during acclimation (Whitehead *et al.* 2012b). Genes associated with water and ion transport systems (aquaporin, kcnj2, v-type proton assembly unit (Whitehead *et al.* 2012b; 2013; Kozak *et al.* 2013; Brennan *et al.* 2015; Moorman *et al.* 2015; Jung *et al.* 2015)) responded to hypo-osmotic shock. Finally, NRDG1 (N-myc downstream regulated) was induced, and has been identified as a key regulator of gill remodeling (Kültz *et al.* 2013). SLC12A2 (NKCC1), which was repressed in all three populations at 24 h, is basolaterally expressed in mitochondrion rich cells, actively transports sodium, potassium and chloride into mitochondrion rich cells from the bloodstream (Brennan *et al.* 2015), and is consistent with a reduced ion export requirements in fish residing in fresh water. The inwardly rectifying potassium channel kcnj1 was also repressed, and plays a key role in teleost osmoregulation (Marshall 2011). Finally, SLC25, an osmotically responsive mitochondrial calcium transporter (Brennan *et al.* 2015), was repressed at 24 h in our populations, as was carbonic anhydrase, which may facilitate chloride secretion in gill ionocytes (Liu *et al.* 2016).

Hypo-osmotic Shock Alters Cellular Resource Management

The common response shared by the three populations extends beyond the genes shown in Figure 2.4 and includes systematic regulation of the functional pathways shown in Figure 2.5. Collectively, patterns of induction and repression at the pathway level are consistent with highly conserved eukaryotic osmotic shock responses as well as specific aspects of response to hypo-osmotic shock.

At 1 hour, genes associated with protein processing in the ER, a pathway that participates in general stress responses (Boyce & Yuan 2006) were significantly biased toward induction, suggesting increased protein translation during osmotic shock (de Nadal & Posas 2015). Cell volume changes, including those induced by hypo-osmotic shock, can interfere with protein trafficking between the ER and the Golgi (Lee & Linstedt 1999) and create ER stress even under stable rates of protein translation. This induction of ER stress genes may therefore be a direct consequence of change in osmolarity (Apodaca 2002). Finally, ER stress is associated with autophagy, which may facilitate tissue remodeling (Mizushima & Komatsu 2011), so ER stress in this system may be a result of re-using existing cellular components as well as fabricating new ones.

Although genes associated with ER stress are systematically induced during acclimation to hypo-osmotic shock in the gill, the overall pattern of response at the level of KEGG pathways is to systematically repress sets of genes associated with KEGG paths. This observation is consistent with altered resource management during stress in which production of stress-related proteins is transiently favored at the expense of proteins associated with growth and proliferation (de Nadal *et al.* 2011).

Modular Networks Facilitate Acclimation

Although KEGG paths are useful, they cannot identify systematic behavior in genes for which no annotation is available, and annotations are based on homology with well-studied model organisms. *F. heteroclitus* and *Danio rerio,* diverged roughly 200 million years ago (Hedges *et al.* 2015), and *D. rerio* is a freshwater fish that lacks the ability to acclimate. Undoubtedly some plastic response seen in *F. heteroclitus* is lineage-specific. I therefore complemented pathway analysis with direct observation of network structure, and found that salinity-responsive genes are significantly more modular than random sets of genes (Figure 2.6) and more tightly controlled than genes in general (Figure 2.7). That the response to hypo-osmotic shock is mediated by networks that are both modular (loosely coupled) and robustly controlled is consistent with our previous work in which killifish were exposed to hyper-osmotic conditions (Shaw *et al.* 2014) and with emerging theories that networks naturally evolve to have fewer connections (Leclerc 2008).

Tightly Controlled Genes Define Geographic Clades

Finally, our data suggest that the quality of tight regulation appears to be conserved itself. As shown in Figure 2.8, genes with the lowest levels of variation, i.e. those in the first quartile for this trait, show similar behavior within, but not between, killifish clades.

Conclusions

Though killifish gene expression responses showed remarkable diversity, consistent with their ability to adapt, selective processes preserve the responses of genes, pathways and network structures, suggesting that network structure may play an indispensable role in stress adaptation.

3. Low Dose Arsenic Reduces Gene Regulatory Network Connectivity and Impairs Phenotypic Plasticity in Killifish

With: Craig Jackson, Dawoon Jung, Stephen P. Glaholt, Celia Y. Chen, Bruce A. Stanton, John K. Colbourne, and Joseph R. Shaw

ABSTRACT: Exposure to arsenic, as sodium arsenite, during osmotic shock reduces the magnitude of the osmotic response, interferes with the structure of co-expression networks, and systematically represses the activation of specific osmotic response pathways. Killifish (N = 36; 4 fish per group) from three populations were acclimated to saltwater for 5 weeks, then transferred to fresh water for 1 h or 24 h. Gene expression profiles were measured in gill tissue from fish exposed to arsenic (100 μg/l) and unexposed fish. Mixed effect linear models were used to estimate the impact of experimental factors on gene expression, namely time in fresh water, exposure to arsenic, and their interactions, with population modeled as a random effect. Hypo-osmotic shock at 1 h or 24 h regulated 31 and 178 genes respectively (FDR < 0.05). In the presence of arsenic, gene expression reaction norms (absolute log2 fold change versus control) were significantly reduced ($p < 2e-6$; paired t test) and the number of genes showing significant co-regulation was reduced at 24 h ($p < 0.001$; permutation test). The presence of arsenic systematically repressed genes in 13 KEGG paths (FDR < 0.05; binomial test). Collectively, these results show that low dose arsenic interferes with the osmotic shock response at the level of individual genes, gene regulatory networks and functional pathways.

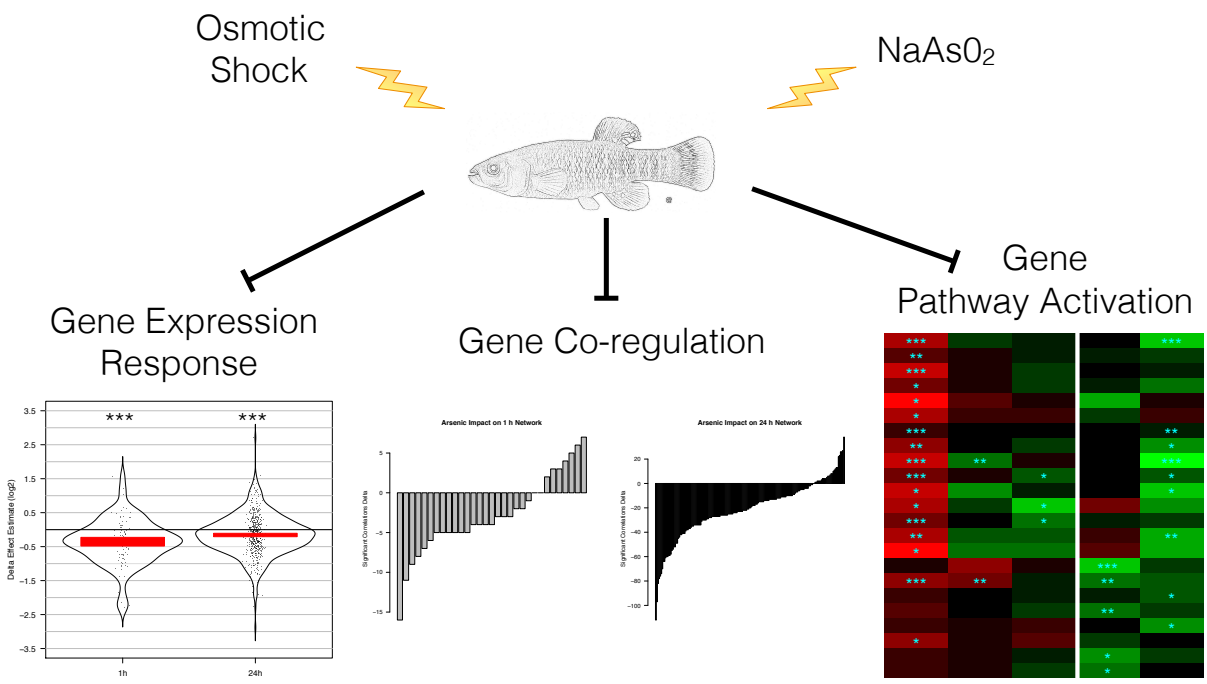

Figure 3-1 Graphical abstract for Environmental Science and Technology

Introduction

A notoriously toxic metalloid, arsenic is naturally abundant in the Earth's crust and is used as a pesticide, pigment, wood preservative and medicine, making it a significant contaminant in many environments (Duker *et al.* 2005). Inorganic arsenic above the part per million range (mg/l or mg/kg) blocks oxidative phosphorylation, leading to acute toxicity and death, but the effects of low dose arsenic are diverse, and not well understood. Trivalent species interact with proteins that contain thiols, and may inactivate as many as 200 enzymes (Ratnaike 2003). Pentavalent species, by contrast, may replace phosphate in a variety of cellular processes (Hughes 2002), altering the behavior of metabolic enzymes and signaling pathways that use phosphate (Rosen *et al.* 2011). Arsenic in the environment is generally below concentrations associated with acute toxicity, yet exposure to environmentally relevant low dose arsenic concentrations, e.g., less than 100 µg/l, has been linked to a growing list of

human diseases involving the skin, nervous system, respiratory system, cardiovascular system, liver, kidney, bladder, immune and endocrine systems (Naujokas *et al.* 2013).

Emerging evidence suggests that the presence of arsenic, even in very small amounts, degrades other organismal responses. For example, mice exposed to low dose arsenic fail to mount an effective response to influenza infection (Kozul *et al.* 2009a), and infants exposed to arsenic in utero show signs of impaired immune response (Farzan *et al.* 2016). Arsenic interferes with wound healing (Olsen *et al.* 2008), response to bacterial pathogens (Goodale *et al.* 2017), and response to increased salinity in the killifish (Stanton *et al.* 2006). The killifish response to osmotic shock is complex and coordinated, involving an early crisis response, in which salinity alteration is sensed and steps are taken to avoid system collapse, as well as long term acclimation that involves dynamic rearrangement of cells in the killifish gill (Marshall *et al.* 1999; Marshall 2003; Stanton *et al.* 2006; Marshall *et al.* 2009; Shaw *et al.* 2010; Whitehead *et al.* 2011b; a; 2012b; Brennan *et al.* 2015). In a recent study (Shaw *et al.* 2014), we used low dose arsenic to block salinity acclimation and observed that arsenic systematically reduces the magnitude of responses to hyper-osmotic shock, suggesting that arsenic degrades responses to other environmental stressors. This behavior, if it generalizes, could explain arsenic's exceptionally broad toxicological footprint. As a first step in testing this hypothesis, I have used our killifish model of arsenic interference to test whether arsenic interferes with gene expression responses to fresh water (hypo-osmotic shock) in the same way that it interferes with gene expression responses to salt water (hyper-osmotic shock).

Our experimental design included three populations of wild killifish (Reid *et al.* 2017), embracing individual variability, and a second stressor (hypo-osmotic shock) to create an environmentally relevant model system. Our choice of arsenic and salinity as co-stressors was based on known toxic interactions for killifish acclimating to osmotic shock at high

doses (Stanton *et al.* 2006), and significant interaction effects between arsenic and osmotic shock at low doses (Shaw *et al.* 2014). In addition, I re-analyzed data from a single population exposed to hyper-osmotic shock to validate that arsenic systematically interferes with both hyper-osmotic and hypo-osmotic responses in similar ways. Specifically, I have looked for statistically significant associations between arsenic exposure and a) the magnitude of response osmotic shock response, b) the strength of correlation is gene networks, and c) the level of activation of genes belonging to certain biological pathways

Collectively, results reported here suggest that arsenic interfere with programmed stress responses at the gene, network and pathway level, possibly contributing to its complex toxicological phenotype (Naujokas *et al.* 2013).

Materials and Methods

Unless otherwise noted, the R programming language was used for statistical tests and data representation (Team 2014b).

Killifish, Exposures, Tissue Collection

Wild killifish were sampled from two locations in Maine (Northern clade) and one from Virginia (Southern clade), as previously described in Chapter 2. Fish were pre-acclimated to common garden conditions, housed, maintained, and exposed to arsenic as previously described (Shaw *et al.* 2014), as further detailed in Supplemental Information. Briefly, 72 killifish were maintained for at least two weeks in seawater, to assure fully acclimation. Half of the individuals from the three killifish populations (Northeast Creek, Horseshoe Creek, King's Creek) were exposed to arsenic (100 μg/l sodium arsenate in swim water) 48 h before the start of the fresh water exposure time course, and throughout the time course. Arsenic

naïve and arsenic-exposed fish were then transferred to fresh water and harvested at 0 h, 1 h, or 24 h to yield a balanced design of 18 groups with 4 fish per group. Harvested fish were anesthetized, pithed, and gills were removed and stored in RNAlater (Qiagen, Valencia, CA) according to manufacturer's recommendations.

RNA Isolation, Hybridization and Normalization

Following harvesting, RNA from homogenized, lysed gill tissue was precipitated in ethanol, and isolated using RNAeasy kits (Qiagen, Valencia, CA). RNA quality was assessed with an Agilent 2100 Bioanalyzer (Agilent Technologies, Wilmington, DE). All samples used in gene expression studies achieved a RNA integrity number (RIN) score >7. RNA was amplified using MessageAmp II kits (Ambion) and hybridized to a custom NimbleGen array that interrogates 135,000 probes associated with 69,426 unique contigs, that map to 16,104 unique genes. Raw fluorescence values were quantile normalized across arrays using RMA (Bolstad & Bolstad 2013).

Identifying Genes That Respond to Hypo-osmotic Shock

When multiple probes mapped to the same killifish gene, the probe with the highest median expression across all conditions was chosen to represent that gene. RMA normalized $\log_2$ expression values for each gene were analyzed by mixed effect linear models using the R nlme package (Pinheiro *et al.* 2015) with two main effects: arsenic, and time spent in seawater, plus all interactions between the main effects, and a single random effect, population. False discovery rates (FDR) were calculated from p values using the method of Benjamini and Yekutieli (Benjamini & Yekutieli 2001), and genes with an FDR less than 0.05 were deemed significant.

Identifying Genes That Respond to Hyper-osmotic Shock

In a previously published experiment (Shaw *et al.* 2014), killifish from a single population were collected and responses to hyper-osmotic shock were measured in experiments that mirrored the experiments described above. Twenty-four killifish acclimated to fresh water were exposed to arsenic (100 μg/l sodium arsenate) during acclimation to saltwater for 0 h, 1 h, or 24 h. Half of these fish were exposed to arsenic throughout the experiment as described above. Linear models were used to detect differential gene expression, defined as a p-value less than 0.05 and a fold change of 2. These genes were used in our re-analysis of these data because the FDR approach in this experiment identified fewer than 10 arsenic-responsive genes, and the fold-change cutoff mechanism recommended by the MAQC (Shi *et al.* 2006; Chen *et al.* 2007) offers a more powerful alternative in such cases.

Quantifying the Impact of Arsenic on the Magnitude of Osmotic Shock Responses

Mean $\log_2$ expression across the four fish in each experimental group was calculated to represent three populations, three time points (0 h, 1 h, 24 h), and two arsenic exposures (0 ug/l, 100 ug/l) resulting in 18 measurements for each gene. Osmotic shock responses were calculated as the absolute difference between the mean expression at 0 h compared to the mean expression at 1 h or 24 h, using values from the same population and arsenic exposure group for these comparisons. The general response across populations was then calculated as the mean response at 1 h or 24 h, and the impact of arsenic on the absolute response at 1 h or 24 h was calculated as the mean response for each gene in arsenic exposed fish minus the mean response in arsenic naïve fish. Visualization of osmotic shock responses delta values for each differentially expressed gene was created using pirate plots in yarrr (Phillips 2017). The significance of the impact of arsenic on the magnitude of osmotic shock response was

estimated using a one sample t test where the expected mean difference under the null hypothesis was zero.

Quantifying the Impact of Arsenic on Gene Co-regulation

Separate correlation matrices were constructed for arsenic exposed and arsenic naïve fish, to explore relationships between genes that respond to osmotic shock as defined above. The number of significant (FDR corrected) gene-gene correlations throughout the network was assessed using the corr.test package in the R psych library (Revelle 2014). The impact of arsenic on the number of significant gene-gene correlations was assessed using a binomial test in which the null hypothesis was that 50% of genes gain significant correlations by chance.

Quantifying the Impact of Arsenic on Pathway Activation

Significant biases toward pathway activation or repression were determined using linear models to identify genes that respond to osmotic shock, as defined above. Killifish genes were mapped to zebrafish genes as defined in "Identifying Well Annotated Genes" in Supplemental Material. Expression values for all genes in all KEGG paths were tested for pathway level biases toward induction or repression using binomial tests, under the null hypothesis that 50% of the genes in any path are induced by chance. Significance levels of binomial tests were FDR-corrected for multiple hypothesis testing based on the 166 KEGG paths that were tested.

Results and Discussion

Arsenic Diminishes Gene Expression Response to Osmotic Shock

Mixed effects linear models identified 31 genes that responded significantly to hypo-osmotic shock at 1 h, and 178 genes that respond to hypo-osmotic shock at 24 h (FDR < 0.05). Differences in the magnitude of response to osmotic shock attributable to the presence of arsenic (response delta values in Figure 3.1) were calculated as the absolute value of the response observed in arsenic-exposed fish less the absolute value of the response observed in fish that were not exposed to arsenic. Genes responding to hyper-osmotic shock were selected as previously described (Shaw *et al.* 2014), and mean absolute values were also compared in the presence and absence of arsenic as above.

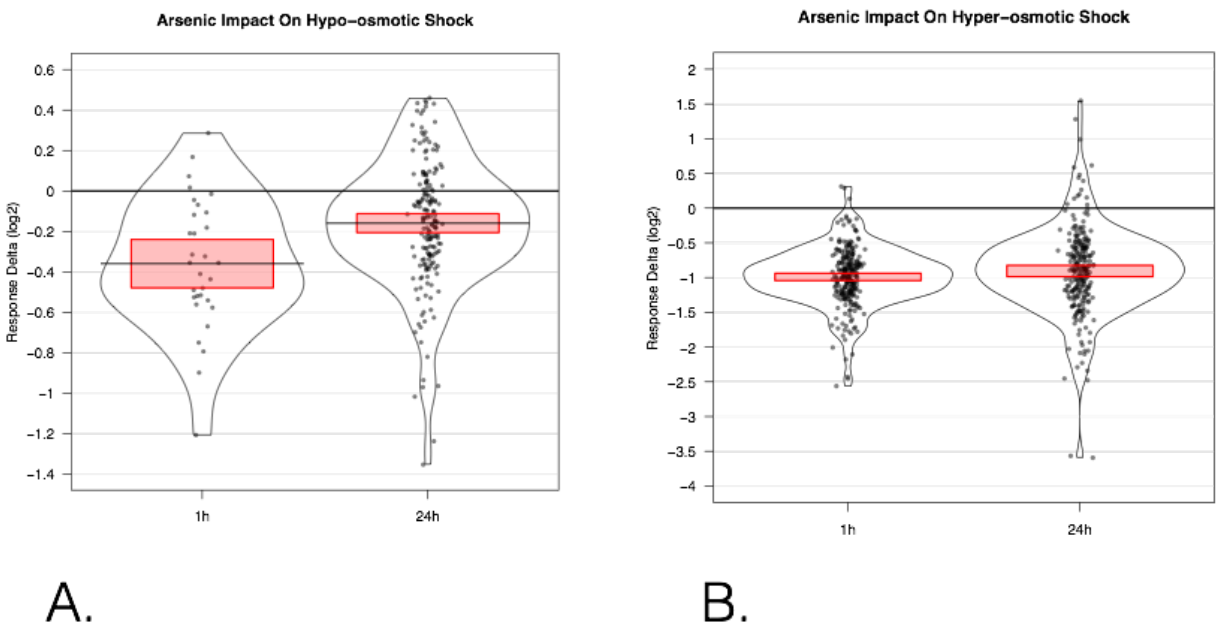

Figure 3-2 Absolute value of responses in arsenic exposed fish less the absolute response in unexposed fish (Response Delta). Each symbol represents a gene that responded significantly at each time point. The frequency of observing a given Response Delta is shown by kernel density estimations. Red bars reflect the 95% confidence interval of the mean delta value (black line). A: The impact of arsenic on hypo-osmotic shock. B: The impact of arsenic on hyper-osmotic shock.

Figure 3-2 shows the impact of arsenic on reaction norms for genes that respond significantly to osmotic shock. The difference in salinity response for each gene is simply the gene expression reaction norm estimated from linear model salinity terms in fish that were exposed to arsenic, less the reaction norm estimates for the same gene measured in fish that were exposed to arsenic.

Arsenic generally repressed responses to hypo-osmotic shock, as shown in Figure 3-2A. At 1 h, gene expression mean responses decreased in the presence of arsenic by -0.36 in log2 units, equivalent to a reduction of 22%. At 24 h, mean response repression attributable to arsenic was 10%. One sample t tests comparing the response of each salinity responsive gene

in the presence or absence of arsenic were significantly less than zero at both 1 h and 24 h (p-value < 2e-6).

Arsenic also repressed response to hyper-osmotic shock (Fig. 1B). Data were re-analyzed from a previously published experiment (Shaw *et al.* 2014), and differentially expressed genes were selected as they were in that publication, namely, a fold change difference greater than 2 and an unadjusted p value less than 0.05. At both 1 h and 24 h arsenic repressed the response to hyper-osmotic by about 50%, and this effect was highly significant ($p < 2.2$ e-16, mean less than zero, one sample t test) as shown in Figure 3-2B.

General interference at two time points and two forms of osmotic shock suggests that arsenic either interferes with a small group of regulatory events, each of which has a large span of control, or it interferes with a large group of regulatory events. It is well understood that acclimation to altered salinity involves both an early, crisis phase, and a later phase in which the more gradual process of gill remodeling occurs, and it has been shown that response to hyper and hypo-osmotic shock involves different sets of genes (Whitehead:2011bo Whitehead *et al.* 2011b; 2012b; Brennan *et al.* 2015). Indeed, Figure 3-2 shows 631 unique genes, of which only 11% significantly responded in more than one condition, suggesting that each time point and response is quite independent. Arsenic has been shown to inhibit kinase signaling cascades (Shim *et al.* 2016), and kinase signaling in LKB1-AMPK signaling pathway (Wang *et al.* 2010), an upstream regulator of TOR, a major regulator of proliferation, motility, transcription and translation. It is therefore tempting to speculate that global depression of transcriptional responses, particularly at 1 h, are at least partly due to arsenic mediated defects in phosphorylation, leading to a muted signaling response. What is known, however, is that arsenic inhibits pathways without decreasing phosphorylation (Liu & Bain 2014; Bain *et al.* 2016), and that arsenic is associated with increased phosphorylation of

kinases of protein tyrosine kinases and their targets (Wetzler *et al.* 2006; Wen *et al.* 2010). Therefore, although inhibition of phosphorylation events may play a role in arsenic-mediated interference in signaling cascades, reduced phosphorylation is not universally observed.

Arsenic Reduces Co-expression of Genes Responding to Osmotic Shock

Co-expression, co-regulation and co-functionality are presumed to be related based on structural considerations and on the understanding that natural selection generally rewards efficiency (Michalak 2008). The prokaryotic genome, for example, is organized into operons containing genes that are functionally related and co-expressed as a group. Co-expression in eukaryotic systems is modulated by many factors other than genomic proximity, including epigenetics, regulatory RNAs and binding sites for regulatory proteins such as transcription factors, but ultimately, co-expression is thought to reflect biological needs (D'haeseleer *et al.* 2000; Segal *et al.* 2003).

In the present study, I observed a global reduction in response to osmotic shock in the presence of low dose arsenic (Figure 3-2), and I wanted to know whether this response also altered co-expression. Reduced co-expression in the presence of arsenic could be construed as evidence of arsenic interfering with co-regulation, or an independent toxic mechanism of action interfering with functionality. In either case, Ayroles and Mackay (Ayroles *et al.* 2009) suggest that the mean absolute value of Pearson r provides a measure of how "connected" genes within a particular gene set are to each other, evoking a common metaphor in which functionally related genes are represented as nodes in a network joined by an edge (Carter *et al.* 2004).

A global picture of connection is provided in Figure 3-3, which plots the frequency with which various levels of correlation were observed in the expression of genes in fish that were

not exposed to arsenic (red lines) and fish that were exposed to arsenic (green lines). The four panels represent distributions of four sets of genes that responded significantly to either hypo-osmotic shock at 1 h or 24h (left) or hyper-osmotic shock at 1 h or 24h (right) as described in Methods. The presence of arsenic systematically reduced correlation in all four gene sets.

Biologically speaking, reduced correlation indicates more independent behavior in expression, which is consistent with loss of control over osmotic responses. Although arsenic consistently shifted correlation toward zero, the correlation distribution at 1 h in hypo-osmotic shock (Figure 3-3A) has a different structure from correlation distribution at 1 h in hyper-osmotic shock (Figure 3-3C). Specifically, Figure 3-3A is dominated by positive correlations, while the other distributions contain both positive and negative correlations. The predominance of positive correlations at 1 h in response to hypo-osmotic shock could indicate that this response is largely mediated by immediate, early response genes (Bahrami & Drabløs 2016), which contribute to population differences in osmotic shock response in killifish (Whitehead *et al.* 2011b).

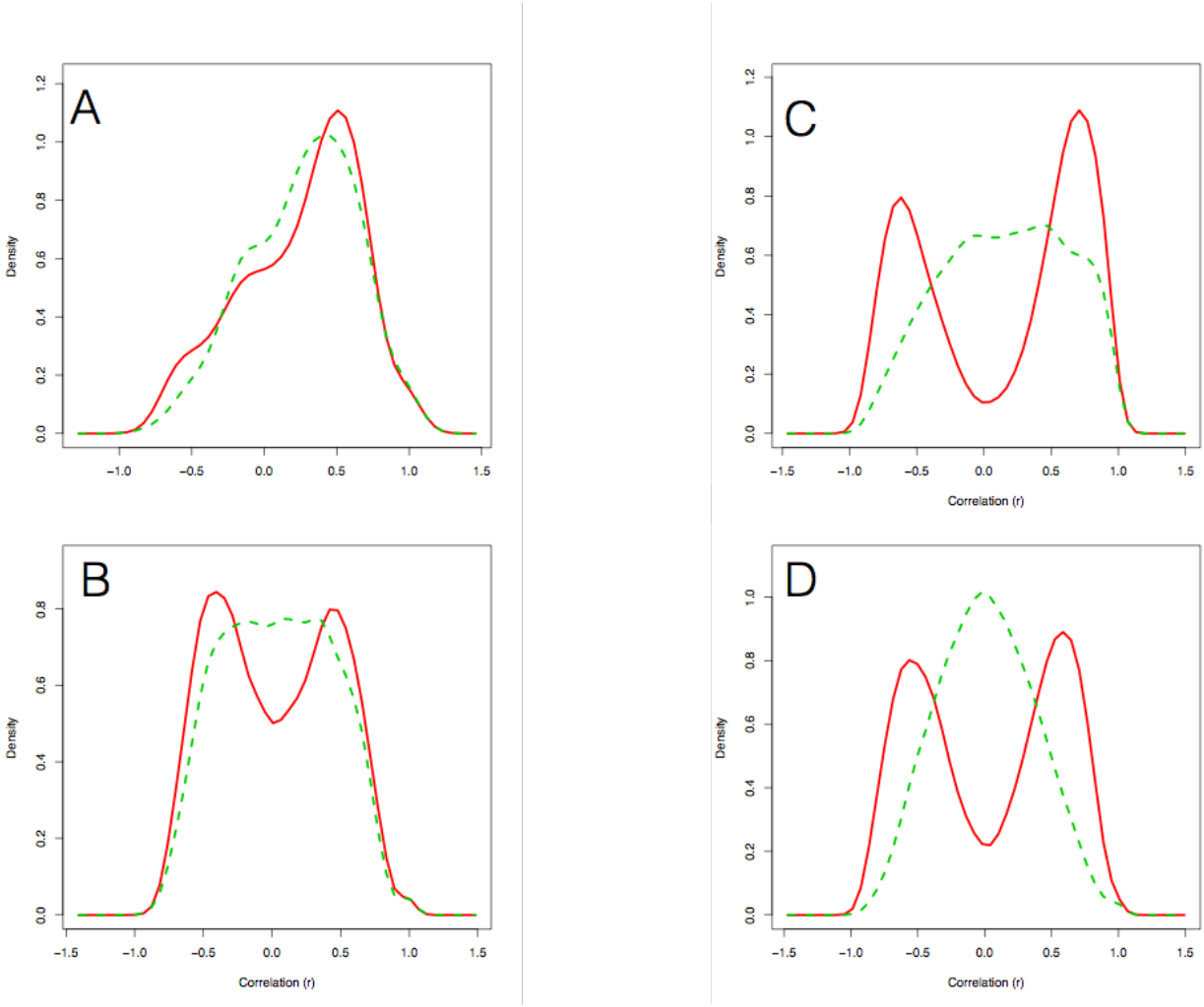

Figure 3-3 Distribution of correlation (Pearson r) values for four sets of genes significantly differentially expressed in response to (A: hypo-osmotic shock, 1 h; B: hypo-osmotic shock, 24 h; C: hyper-osmotic shock, 1 h; D: hyper-osmotic shock, 24 h). Solid red lines indicate Pearson r value distribution for samples that were not exposed to arsenic. Dashed green lines indicate arsenic exposed samples.

Figure 3-3 reveals differences in the distribution of correlation, but I was interested in identifying which genes experience the greatest, most meaningful changes. To explore this, I began by making a distinction between gene-gene correlations that achieve significance (FDR < 0.05) and those that do not, under the assumption that statistically significant correlations are most likely to be biologically relevant. Next, for each gene in any set of regulated genes, I compared the significant correlations among genes in the set in the

presence of arsenic to the number of significant correlations in the absence of arsenic. This yields the sorted distribution of gains and losses in significant correlations attributable to arsenic shown in Figure 3-4. Delta values are significantly skewed toward loss in every case ($p < 0.009$, binomial test), but by no means uniform: arsenic appears to target some genes preferentially.

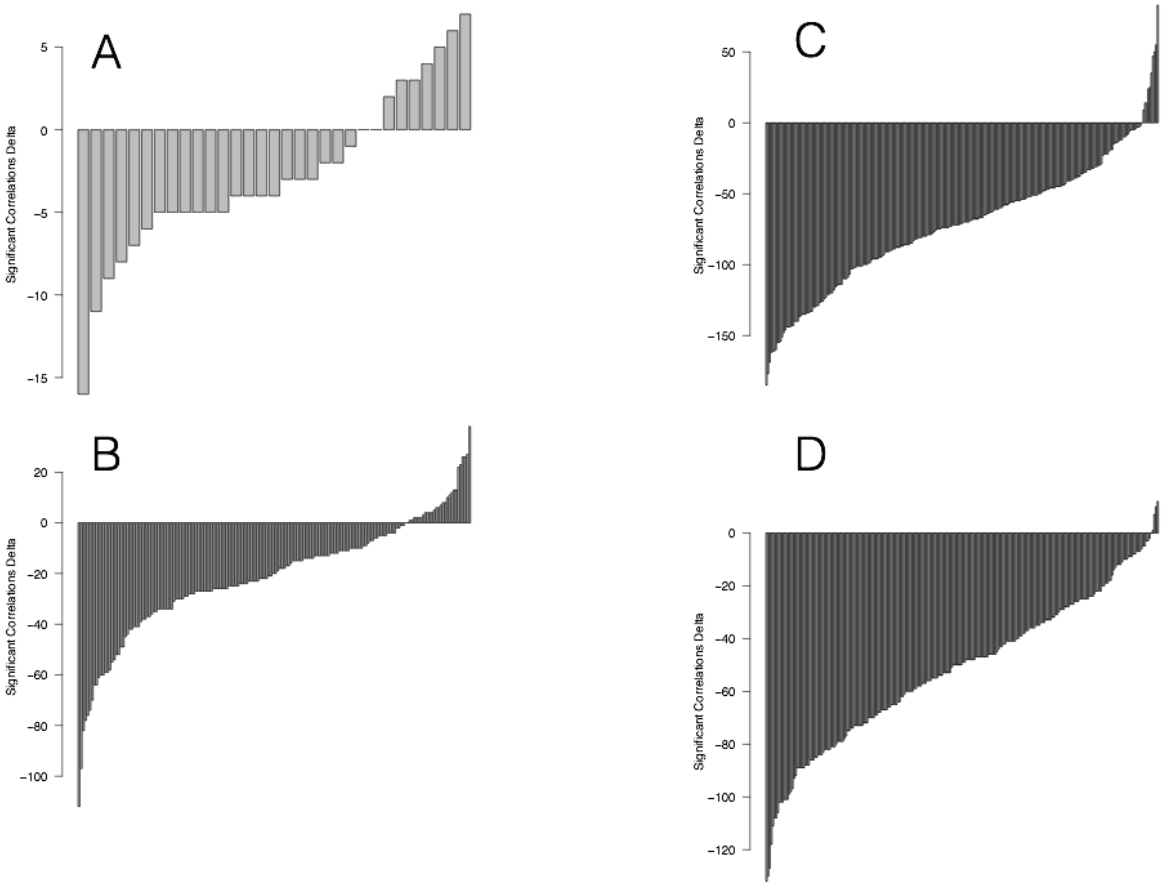

Figure 3-4 Net gain or loss in significant correlation for salinity responsive genes in arsenic exposed or arsenic unexposed fish during osmotic shock. A: hypo-osmotic shock, 1 h; B: hypo-osmotic shock, 24 h; C: hyper-osmotic shock, 1 h; D: hyper-osmotic shock, 24 h.

Several recent reports suggest that loss of co-expression in the presence of low dose arsenic may explain loss of ability to acclimate to arsenic at higher doses(Stanton *et al.* 2006). Co-expression networks are conserved (Mähler *et al.* 2017), loss of co-expression is associated with adverse responses to multiple stressors (Pang *et al.* 2017), and loss co-expression is a hallmark of signaling in cancer networks (Anglani *et al.* 2014). On the other hand, increased coordination between cadmium-responsive genes is a hallmark of cadmium tolerance (Nota *et al.* 2013).

Transcription Factors Significantly Induced by Hypo-Osmotic Shock

Above, I have shown that arsenic reduces the magnitude of the osmotic shock response and reduces correlation between genes, and that loss of correlation is more pronounced in some genes as opposed to others. In this section, I explore which genes are most prone to lose significant correlations in the presence of arsenic, focusing only on the hypo-osmotic shock response, the new data presented in this report.

We began by hypothesizing that the simple, one-peaked distribution of the 1 h response to hypo-osmotic shock (Figure 3-3A) might be explained by relationships between transcriptional activators and their targets, many of which are significantly induced, as shown in Figure 3-5. The response to hypo-osmotic shock alone is shown by red bars, and the combined impact of arsenic and hypo-osmotic shock is shown in blue. The biological interpretation of Figure 3-5 is that arsenic does not significantly interfere with the expression

response of transcription factors, with the exception of Hif1a ($p < 0.001$, mixed effect linear model).

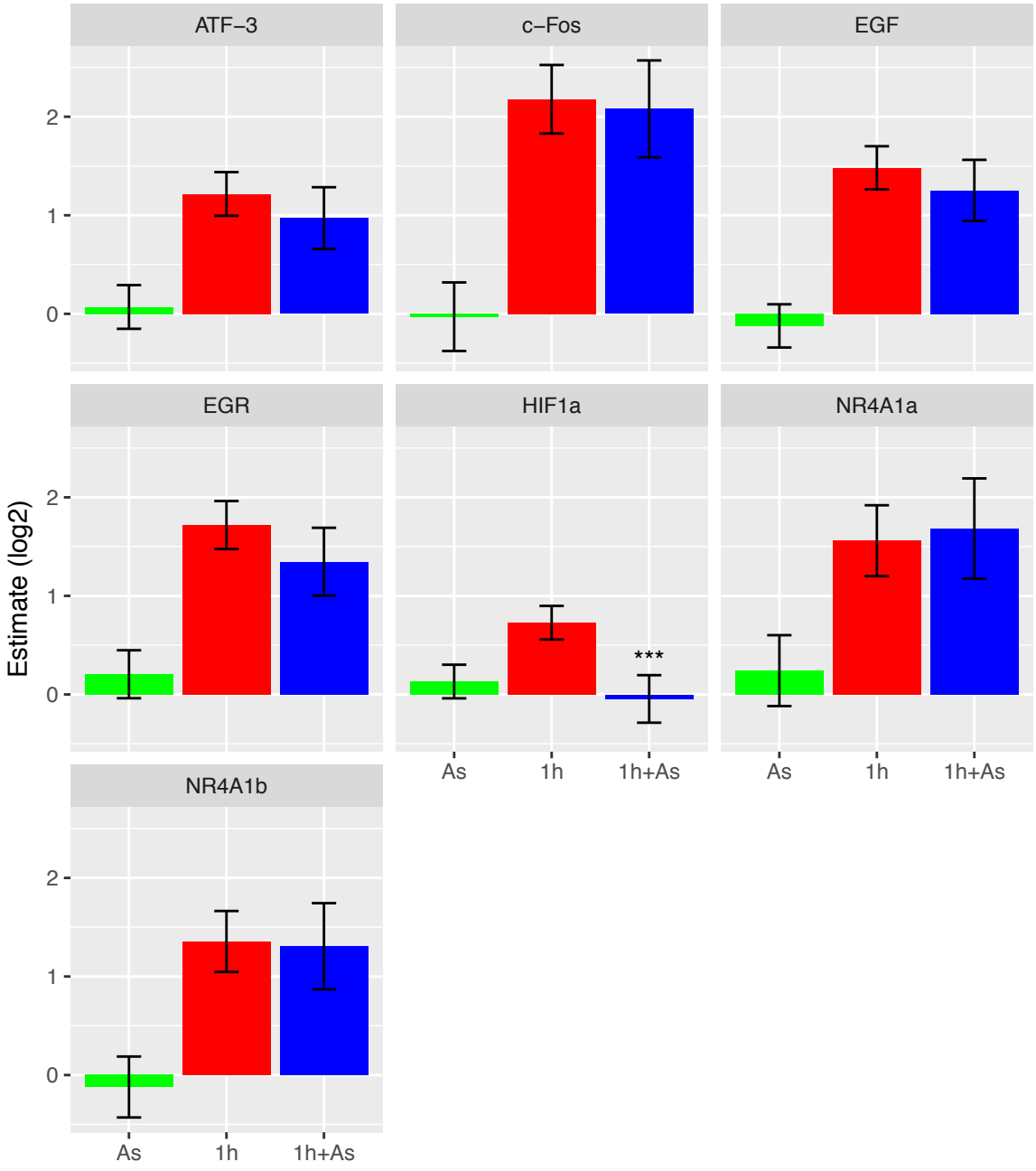

Figure 3-5 Transcription factor responses to arsenic (As), 1 h exposure to hypo-osmotic shock, and the combination of arsenic and 1 h exposure to hypo-osmotic shock (1h+As), in $\log_2$ units relative to control. Error bars are 95% confidence intervals.

Another simple model of how arsenic might interfere with gene expression responses during hypo-osmotic shock would be for arsenic to target signal transduction relationships between upstream regulators, such as transcription factors, and their downstream targets. In this model, one expects the upstream regulators to be induced, even when arsenic is present, but for their targets to respond less than they otherwise would. As I will demonstrate, our data support this model for three reasons. First, the presence of arsenic reduces co-regulatory relationships between transcription factors and other differentially expressed genes at 1 h. Second, arsenic reduces co-regulatory relationships at 24 h, and genes that lose the greatest number of significant co-regulatory relationships show the greatest loss in transcriptional response. Third, arsenic's interaction effect on pathways is inhibitory, which is consistent with interference with signal transduction.

Arsenic Reduces Transcription Factor Co-regulatory Relationships at 1 h

Table 3-1 shows the same genes as those shown in Figure 3-4A, genes that responded to hypo-osmotic shock at 1 h (FDR < 0.05). Here I am focusing on the 1 h response, because this provides a measure of the dynamic response to acclimation, i.e., it is the period of greatest change. Where Figure 3-4 shows that the presence of arsenic rearranges co-regulation, Table 3-1 is sorted by the net number of significant co-regulation relationships that are gained or lost ("Delta") due to arsenic. The transcription factors shown in Figure 3-5 are highlighted in Table 3-1, demonstrating that exposure to arsenic reduced the correlation between salinity-responsive transcription factors and other salinity responsive genes.

Significant Gene	Delta
Uncharacterized protein	-16
Coagulation factor II (thrombin) receptor 1	-11
Early growth response protein (EGR)	-9
Endothelial lipase	-8
Hypoxia inducible factor 1, alpha subunit (basic helix-loop-helix transcription factor) (HIF1a)	-7
Sec23 B	-6
Cysteine/serine-rich nuclear protein	-5
IGF family receptor	-5
Cyclic AMP-dependent transcription factor ATF-3 (ATF3)	-5
Nuclear receptor subfamily 4, group A, member 1 (NR4A1a)	-5
Proheparin-binding EGF growth factor (EGF)	-5
Ephrin-A4	-5
ADP-ribosylation factor protein 5C	-4
Type III iodothyronine deiodinase	-4
Endothelial differentiation-related factor	-4
Proto-oncogene c-Fos (c-Fos)	-4
Glutamine gamma-glutamyltransferase	-3
Growth arrest and DNA damage-inducible protein GADD45 beta	-3
Nuclear receptor subfamily 4, group A, member 1	-3
Regulator of G-protein signaling	-2
YrdC domain containing (E, coli)	-2
Uncharacterized protein	-1
Gap junction beta-3 protein	0
CCAAT/enhancer-binding protein beta	0
Zinc finger protein 36, C3H type, (mouse)	2
FK506-binding protein	3
Lipid phosphate phosphohydrolase	3
Tensin-4	4
Arginase, type II	5
Thrombospondin type-1 domain-containing protein 4	6
Epithelial mitogen (mouse)	7

Table 3-1 1h hypo-osmotic shock genes, arsenic impact on gene-gene correlations (Delta)

Co-regulation loss at 24 h Associated with Reduced Response

Figure 3-5 shows that exposure to hypo-osmotic shock induces the expression of many immediate early response transcription factors, consistent with previous reports in killifish. Interestingly, arsenic does not significantly interfere with their induction, but, as shown in Table 3-1, arsenic disconnects these genes from other genes significant at 1h. Certainly, disconnecting transcription factors from their downstream targets at 1 h might be expected to lead to reduced responses or co-expression at 24 h. Indeed, genes that lose the more significant correlation lose the largest responses: Differences in correlation attributable to arsenic predict about 17% differences in response attributable to arsenic and the two are significantly correlated (p-value: 4.31e-09), as shown in Figure 3-6.

The biological interpretation of these results is that arsenic exposure may lead to a modest repression of many cellular responses, perhaps by interfering with and degrading gene regulatory networks in a subtle, but pervasive way. Individual examples of this have been observed at higher doses, such as the destabilization of Gli transcripton factors in response to arsenic trioxide (Kim *et al.* 2010), inhibition of JAK tyrosine kinase (Cheng *et al.* 2004), and inhibition of the phosphorylation of protein kinase B (Paul *et al.* 2007). Arsenic may also act as an endocrine disruptor (De Coster & van Larebeke 2012). Arsenic is known to interact with over 300 proteins (Zhang *et al.* 2015), and it is not surprising that exposure to arsenic elicits many distinct effects.

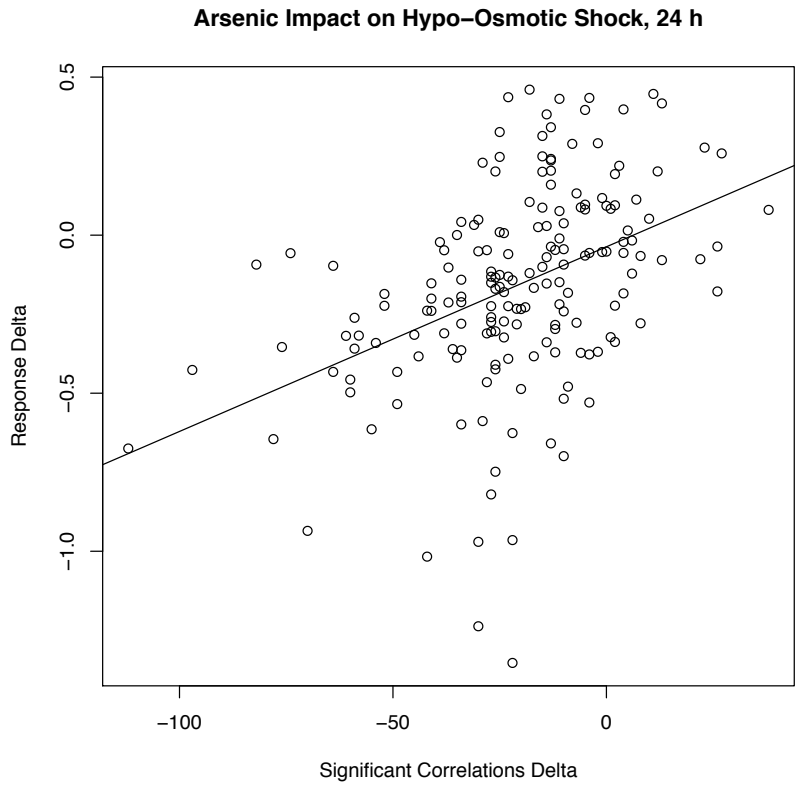

Figure 3-6 Relationship between arsenic impact on number of significant correlations (x axis) and arsenic impact on magnitude of gene expression response (y axis). Each symbol represents a single gene that significantly responded to hypo-osmotic shock at 24 h. Simple linear regression line shown.

Arsenic Interactions Inhibit Pathway Responses at 1 h

Exposure to high levels of arsenic, about 100 times the concentration used in this experiment, has been shown to cause mortality in killifish as they transition from fresh water to salt water (Stanton *et al.* 2006), and I predicted that exposure to much lower concentrations of arsenic (100 μg/l) would be accompanied by observable transcriptomic differences in groups of

functionally related genes that might explain this. Complete pathway results are summarized in Figure 3-7.

Systematic activation of more than 17 pathways by arsenic is shown by red colors in the first column of Figure 6. This is consistent with induction of MAPK signaling, activation of AP-1 and c-fos (Hughes *et al.* 2011), toxic response (Bhattacharya *et al.* 2007), and inflammatory signaling (Farzan *et al.* 2017). Far fewer pathways responded to osmotic shock than arsenic, 2 paths at 1 h and 3 paths at 24 h. Pathway activation values for 1 h hypo-osmotic shock were significantly inversely correlated with 1 h arsenic interaction effects ($r = -0.80$, $p = 5.0 e -6$) but activation values at 24 h were slightly positively correlated with 24 h arsenic interaction effects ($r = 0.38$, $p = 0.28$). This suggests that arsenic interference at the pathway level is most acute during the earliest phases of acclimation, when changing salinity is sensed, and early stress responses are initiated through multiple signaling cascades (Kültz 2012). Notably, induction of protein response in the endoplasmic reticulum may signal rapid protein translation at 1 h in arsenic unexposed fish, as rapid translation is a core stress response in eukaryotes (Lu *et al.* 2004). Significant antagonistic responses in the ER stress response at 1h suggest a delayed or deficient translational response. This is consistent with the phenotype of failed acclimation to salinity in the presence of arsenic, reduced transcriptional responses, and interference with gene regulatory networks.

We believe that these arsenic effects, observed during osmotic shock in the killifish, illustrate the importance of measuring toxicological endpoints in dynamic systems rather than in static

environments, and further, that future research into arsenic's ability to interfere with stress responses may lead to an improved understanding of this complex toxicant.

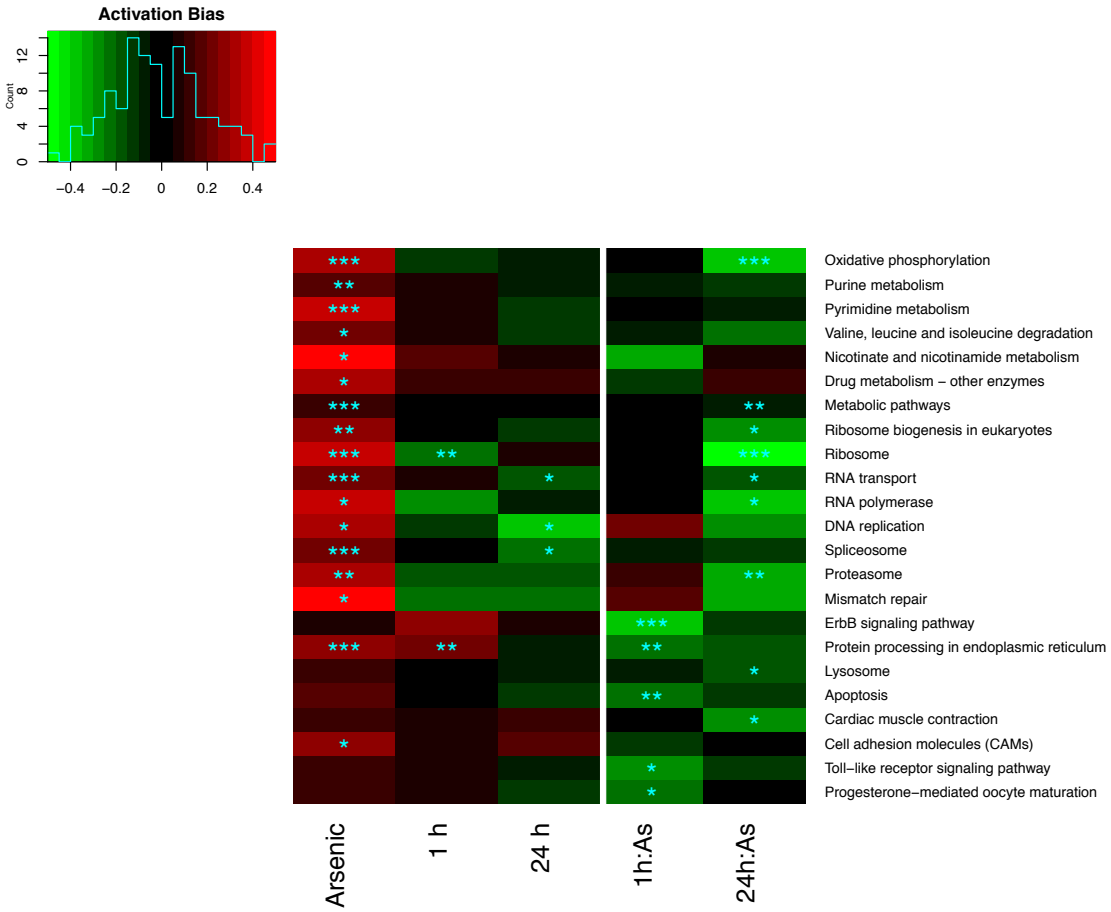

Figure 3-7 Pathway activation analysis of arsenic, hypo-osmotic shock at 1 h, 24 h, and interaction effects between osmotic shock and arsenic exposure. Cell colors indicate systematic bias toward pathway induction (red) or repression (green) and FDR corrected bias significance based on binomial tests is indicated by asterisks ("*" = 0.05, "**" = 0.01, "***" = 0.001). Interactions between arsenic and osmotic shock appear on the right.

4. Enhanced Gene Expression Response to Cadmium Stress May Facilitate Cadmium Tolerance in *D. pulicaria*

With: Craig Jackson, Barbara Pietrzak, Dörthe Becker, Dagmar Frisch, Marcin W. Wojewodzic, Stephen P. Glaholt, Celia Y. Chen, Bruce A. Stanton, John K. Colbourne, and Joseph R. Shaw

ABSTRACT: *Daphnia* populations downwind of smelting operations in Sudbury, Canada, exhibit enhanced tolerance to cadmium compared to *Daphnia* from pristine areas. Transcription profiles were collected to compare cadmium-sensitive and cadmium-tolerant clones at baseline and in response to cadmium. Gene expression responses to five other toxicants, to which the two cones are equally sensitive, were also measured. 44% of *Daphnia* genes differed between the two clones in their baseline expression (FDR < 0.05) compared to 1% that responded significantly in either clone to cadmium exposure. Although sensitive and tolerant clones responded to cadmium stress using the same paths (e.g., glutathione metabolism and cytochrome P450 drug metabolism) and cadmium gene expression responses were correlated between the two clones (Pearson correlation = 0.8), the tolerant population exhibited 30% larger gene expression responses to cadmium ($p < 0.0005$) on average, and its network structure was 29% enriched in significant gene-gene correlations ($p < 0.001$). Increased response to cadmium and increased number of gene interactions were specific to cadmium, and not observed in response to arsenic, atrazine, acetaminophen, nickel and copper. Collectively, our results suggest that cadmium tolerance in *Daphnia* is facilitated by systematically larger, more coordinated gene expression responses to cadmium, consistent with cadmium tolerance in *Folsomia candida*.

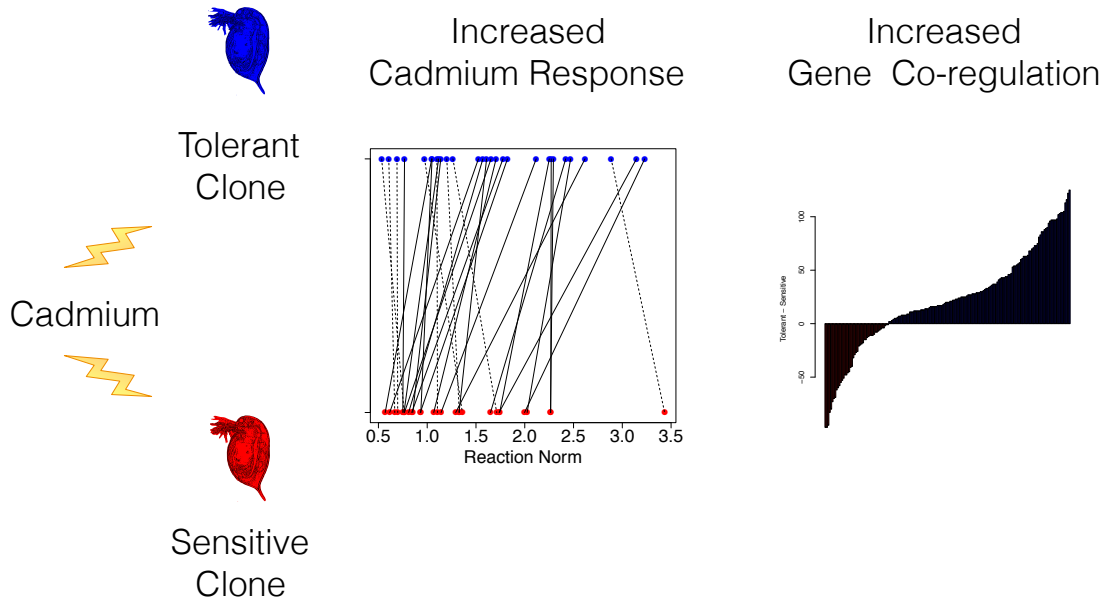

Figure 4-1 Graphical abstract for Environmental Science and Toxicology.

Introduction

Bacteria, plants and animals adapt to antibiotics, herbicides, pesticides and anthropogenic pollution, using a variety of molecular mechanisms. Heritable changes that facilitate tolerance often result from genetic changes such as alternative coding sequences, copy number variation or changes in regulatory sequences (Palumbi 2001; Li *et al.* 2006; Davies & Davies 2010; Wirgin *et al.* 2011; Whitehead 2014; Reid *et al.* 2016). A better understanding of how populations adapt to environmental change and become tolerant to environmental stress informs climate change decisions, guides antibiotic and pesticide strategies, and leads to improved environmental policies.

In the simplest case, resistance to a toxic stress can evolve based on changes in a single gene. For example, houseflies developed resistance to DDT through the insertion of a mobile element into the promoter region of cyp6G1, leading to overexpression of this cytochrome

p450, enhanced ability to metabolize DDT, and therefore DDT resistance (van Straalen *et al.* 2011). Tolerance may also be facilitated by changes in gene regulatory networks. For example, cadmium tolerance in *Folsomia candida* clones (Nota *et al.* 2013), parthenogenetically reproducing hexapods, has been attributed to systematically larger gene expression responses to cadmium (a larger number of genes, and larger fold changes) and gene regulatory networks with higher levels of co-regulation.

Daphnia are freshwater crustaceans that are sensitive to biotic and abiotic stressors, making them useful animal sentinels of environmental stress (Colbourne *et al.* 2011; Kim *et al.* 2015). In this study, a cadmium tolerant *D. pulicaria* clone from a lake downwind of smelting operations in Sudbury, Canada, was compared to a cadmium sensitive clone from a similar lake in Dorset, Canada, that was unaffected by smelting operations. Hereafter, the two clones are referred to as the Cd-tolerant and Cd-sensitive clones, respectively. Although the two *D. pulicaria* clones used in this study differ in their sensitivity to cadmium, they are equally sensitive to arsenic, atrazine, acetaminophen, nickel and copper, based on tests of reproductive fitness following chronic exposure. Cadmium gene expression responses that might explain cadmium tolerance differences are therefore those that a) differ between the two clones and b) are uniquely associated with the cadmium response. Gene expression response differences meeting these criteria were identified at the level of individual genes, but I also identified systematic effects involving many genes such as biological pathways, global magnitude of response, and gene regulatory network structure. I found that many of the same genes and pathways were induced in response to cadmium stress in both tolerant and sensitive clones, but the tolerant clones induced these genes and pathways to a greater extent, suggesting that enhanced gene expression response facilitates tolerance to environmental stress.

Materials and Methods

Animals

Daphnia pulicaria used in this study were obtained from laboratory cultures originally collected from McCharles Lake, Sudbury, Ontario (Cd-tolerant) and from Glen Lake in Dorset, Ontario, Canada (Cd-sensitive). *Daphnia* were housed in 1 L borosilicate glass beakers (20 per beaker) held inside an environmental chamber at a constant temperature (20 +/- 1^{o} C) and photoperiod (16:8 h light:dark). Organisms were maintained in nanopure water reconstituted to moderate hardness (COMBO media (Kilham *et al.* 1998)) and renewed weekly. *Daphnia* were fed daily with *Ankistrodesmus falcatus* to provide 1.5 mg C L^{-1} of culture media. Maternal effects were controlled for and all individuals used in the experiments were guaranteed to be the same age during experiments by isolating neonates (<24 h old) from maintenance cultures one generation before the initiation of a toxicity test. These organisms, referred to as brood females, were synchronized with respect to age and provided neonates for tolerance assessment and microarray experiments as described below.

Tolerance Assessment

We determined tolerance levels of the two clones to each of the six stressors using chronic 21 d lifetable experiments as described in Asselman et al. (Asselman *et al.* 2012). For each experiment, isolates were individually housed in 30 mL of COMBO media, and ten clonal replicates were exposed to each of five test concentrations and control conditions. Total reproduction was assessed over 21 days and the concentration that inhibited reproduction by 25% was calculated (IC25). Only the cadmium elicited statistically different results between clones ($p<0.029$), with the tolerant clone exhibiting significantly better reproductive success.

Toxicant Exposures

Daphnia were acutely exposed (48 h) to LC10 concentrations of toxicants determined for the Dorset clone in two batches, performed in two consecutive years. The LC10 doses were as follows: arsenic: 1,000 μg/l, atrazine: 100 μg/l, acetaminophen 25 μg/l, nickel 200 μg/l, cadmium 20 μg/l, and copper 1 μg/l. First year exposures included arsenic, acetaminophen and atrazine, arsenic + acetaminophen, arsenic + atrazine, and acetaminophen + atrazine. Second year exposures included cadmium, copper, nickel, cadmium + copper, cadmium + nickel, and nickel + copper. Exposures included ten 15-day old *Daphnia* housed in 1 L beakers. For each stressor, stressor combination or control (COMBO media) four biological replicates were exposed for 48 h and RNA was immediately extracted.

RNA Isolation and Extraction

RNA extraction and DNAse treatment were performed on fresh tissue with the Qiashredder and RNeasy columns (Qiagen) following manufacturer's protocol. The quality and quantity of RNA were assessed with a Nanodrop spectrophotometer and Bioanalyzer (Agilent), respectively, as described in Shaw et al. (Shaw *et al.* 2007). Afterwards, RNA was aliquoted and stored at −80°C until its use in the microarray experiments.

Microarry Processing

The microarray experiment was conducted as described by Lopez and Colbourne (Lopez & Colbourne 2011). One microgram of total RNA was amplified using the MessageAmp II aRNA Amplification kit (Ambion, Applied Biosystems, Carlsbad, CA, USA) following manufacturer's protocol. Double stranded cDNA was synthetized with SuperScript Double-Stranded cDNA Synthesis Kit (Invitrogen, Carlsbad, CA, USA) following an alkaline

hydrolysis clean up (Qiaquick columns, Qiagen). Samples were labeled with Dual-Color DNA Labeling Kit (Roche Nimblegen, Madison, WI, USA) following manufacturers protocol. Following each step (amplification, double strand synthesis, and labelling), quantity and quality of the samples were determined with the spectrophotometer and Bioanalyzer 2100.

The microarray design included a total of 56 samples (7 treatments per clone, each represented by four biological replicates) which were hybridized to five 12-plex NimbleGen arrays using a loop design that included four biological replicates per treatment with dye-flips, as shown in Supplemental Figure 4-7. Treatments for year 1 included atrazine (100 μg/l), acetaminophen (25 μg/l), and arsenic (1,000 μg/l), and for year two cadmium (20 μg/l), copper (1 μg/l), and nickel (200 μg/l). Labeled samples were pooled according to the design, dried, and resuspended in hybridization buffer according to Roche NimbleGen's User Guide for Expression Analysis for Cy-labeled cDNA derived from Eukaryote systems. Subsequent hybridization of each of these pools on the respective arrays followed the same protocol (Lopez & Colbourne 2011) and was executed with the NimbleGen Hybridization Kit (Roche Nimblegen, Madison, WI, USA). Following hybridization, the slides were washed with NimbleGen Hybridization Wash Buffers (Roche Nimblegen, Madison, WI, USA). The microarray was a transcriptome array developed by the Centre for Genomics and Bioinformatics (Indiana University, Bloomington, IN, USA), with its design deposited in the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus under the accession number (GEO: GPL11278). Following hybridization, the arrays were scanned with the NimbleGen MS 200 Microarray Scanner to measure fluorescence and images were processed with NimbleScan 2.6 Software.

Gene Expression

Gene expression values for each sample were inferred from the loop design using linear models, quantile normalized, log2 transformed, and technical replicates were averaged. Genes for which there were multiple replicate probes were represented by the probe having the highest median expression across all samples.

Identifying Differentially Expressed Genes

Baseline expression differences between Cd-sensitive and Cd-tolerant clones were estimated by mixed effect linear models of expression in control samples from the two exposure years, modeling clone type as the main effect and year as a random effect. Treatment effects were estimated using four separate linear models, modeling each toxicant or binary combination of toxicants as a categorical independent variable using control samples from the same clone and year as the reference condition. Genes with FDR (Benjamini & Hochberg 1995) corrected p-values less than 0.05 were considered significant.

Identifying Systematic Differences in Response Magnitude

Response magnitude was defined as the absolute value of the $\log_2$ treatment response relative to the references condition, namely, the unexposed samples from the same year in the same clone. T tests were used to assess the significance of systematic differences in response magnitude: specifically, Welch's t tests were used for unpaired comparisons and paired t tests when the expression of the same genes were compared between clones.

Pathway Gene Set Enrichment and Activation

We identified KEGG (Kanehisa & Goto 2000; Kanehisa *et al.* 2007) pathways that were significantly targeted by cadmium using Fisher's Exact tests where the null hypothesis was

that there was no association between a gene being significantly being differentially expressed and being on a path. The alternate hypothesis was that the association between being a gene that was significant and being on a specific path was greater than expected by chance. Significant pathway activation was measured using a binomial test in which the null hypothesis was that the probability of a significant gene in a path was also induced was 0.5.

Gene Regulatory Network Effects

The 177 genes that significantly responded to cadmium (FDR < 0.05) were used to construct a correlation matrix based on 28 measurements of each gene in a given clone in Year 2, namely, 4 biological replicates of each of 7 treatments (control, cadmium, nickel, copper, and binary combinations). The significance of the pairwise correlations between the 177 genes was scored using the corr.test function in the R psych package (Revelle 2014). Log ratios were calculated between the number of significant correlations in the tolerant and sensitive clones, and systematic differences were assessed using a paired t test.

Statistics

Unless otherwise noted, statistical tests, figures, and other calculations were performed using the R base package (Team 2014b). Multiple hypothesis corrections for false discovery rates (FDR) were calculated using the method of Bejamini and Hochberg (Benjamini & Hochberg 1995).

Results and Discussion

Sensitive and Tolerant Clones Differ at Baseline

10,931 genes (44%) were significantly different (FDR < 0.05) between clones that were unexposed to any toxicant based on mixed effect linear models, and many significant genes (above the dashed horizontal line in Figure 4-2) had a $\log_2$ fold change greater than 4 (Figure 4-2, x axis).

Many detoxification genes including cytochrome P450, glutathione, superoxide dismutases, and mutidrug efflux pumps were constitutively expressed at higher levels in the tolerant clone (Supplemental Table 1) raising the possibility that cadmium tolerance in these *Daphnia* are a function of being "prepared" to handle toxic stress. However, different genes from the same categories were constitutively expressed at higher levels in the sensitive clone than the tolerant clone, making it hard to argue that the tolerant clone is, in general, better prepared to handle toxic stress than the sensitive one. I therefore concluded that factors other than baseline differences in gene expression of individual, well-established toxic response genes contribute to cadmium tolerance, and explored pathway level differences in constitutive expression.

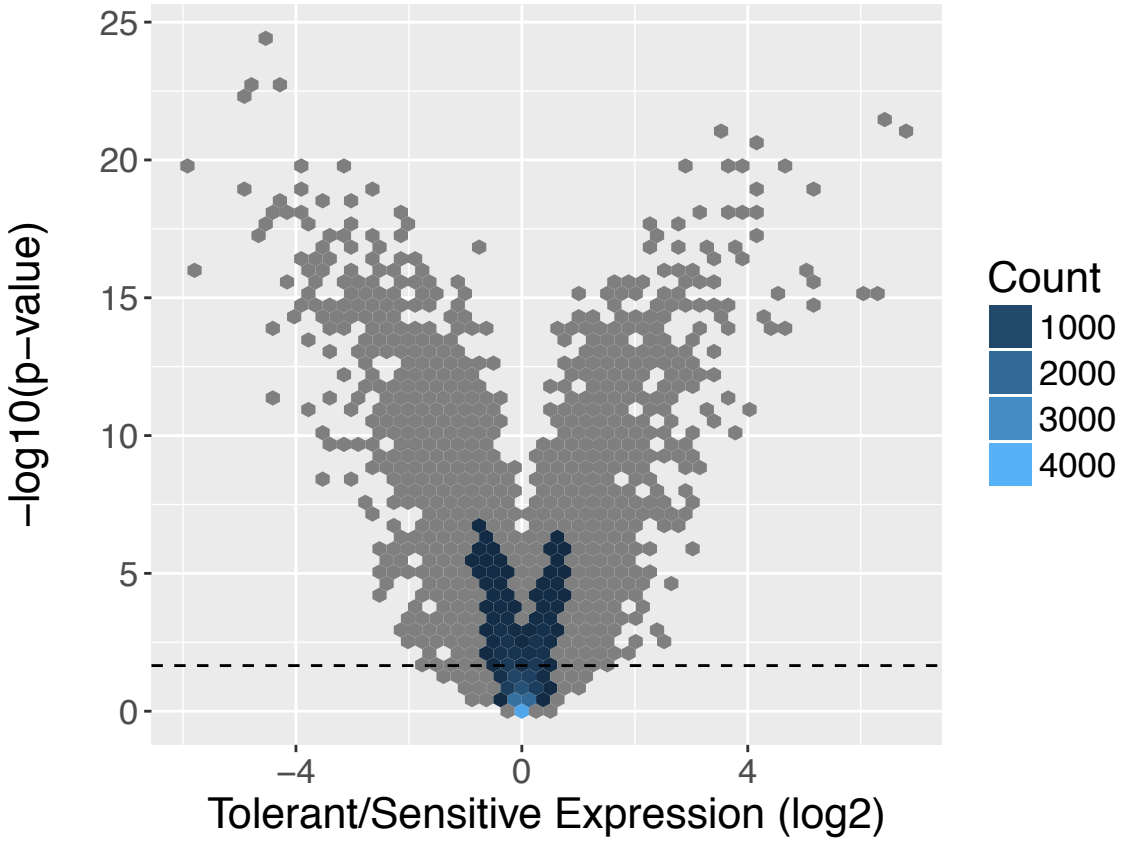

Figure 4-2 Baseline population expression differences. Gene expression ratios (log_2) between tolerant and sensitive clones in control conditions, plotted as a function of significance ($-log_{10}$ p-value). Genes above the dashed horizontal line achieved an FDR of 0.05. Hexagons colors denote number of observations falling within each area.

Glutathione Metabolism Pathway Repressed at Baseline in Tolerant Clone

Three KEGG paths were significantly enriched (FDR < 0.05) in genes that differed in their baseline expression between clones, as shown in Table 1. All three paths are part of a toxic response and show a bias toward repression (less than 50% activation) in the tolerant clone. Based on glutathione's protective role as an anti-oxidant, its function in phase II detoxification, and its ability to chelate cadmium (Li *et al.* 2006; Sanjaya *et al.* 2008; Nguyen *et al.* 2014; Jozefczak *et al.* 2014; Pedrosa *et al.* 2017), it is surprising that baseline gene expression of this path indicates reduced glutathione metabolism in the tolerant clone.

Therefore, I conclude that pathway level differences in constitutive gene expression do not explain the phenotype of increased tolerance, leading me to explore differences in cadmium response.

KEGG Path	Genes	Enrichment	p-value	Activation	p-value
Glutathione metabolism	81	2.14	5.42E-04	0.35	4.89E-02
Metabolism of xenobiotics by cytochrome P450	47	2.69	8.45E-04	0.44	5.97E-01
Drug metabolism - cytochrome P450	45	2.79	7.45E-04	0.48	1.00E+00

Table 4-1 Tolerant vs Sensitive at Baseline: Pathway enrichment and activation of three KEGG paths that were significantly enriched (FDR < 0.05) in genes that differed between the two clones in their baseline expression. The number of genes on the path, the enrichment rate, enrichment p value, activation rate, and activation p value is shown. Activation rate specifies the fraction of significant genes that were induced in the tolerant clone relative to the sensitive clone, and the activation p value is based on a binomial test in which the null hypothesis is that the activation rate is 0.5.

Cd-Sensitive and Cd-Tolerant Clones Share Many Pathway Responses to Cadmium

Cadmium responses in both clones are consistent with toxicant exposure (Bucheli & Fent 2009) in general and metal exposure in particular (Ki *et al.* 2009; Chen *et al.* 2016b). Eight genes with annotations related to glutathione or multidrug resistance were induced by a $\log_2$ fold change of more than 0.5 in both clones. Eight other genes associated with glutathione and multidrug resistance as well as a metallothionein gene (MT-2) were induced by $\log_2$ greater than 0.5, only in the sensitive clone (Supplemental Table 2). No genes with known annotations relating to metallothionein, glutathione, MTF-1, multidrug resistance, cytochrome P450, superoxide dismutase, or heat shock were strongly induced only in the tolerant cone. It should be noted that the Cd-regulated metallothionein gene was excluded from the microarray, because it is duplicated in the reference genome so unique probes could not be designed. At the pathway level, glutathione metabolism and cytochrome P450

pathways were significantly enriched in differentially expressed genes in both clones, as shown in Tables 2, 3 respectively. Enrichment of pentose and glucaronate interconversion path was consistent with the hypotheses that cadmium exposure changes redox balance (Connon *et al.* 2008; Rand *et al.* 2015; Stincone *et al.* 2015). Arachidonic acid metabolism, a biomarker of metal exposure in *Daphnia* (Antczak *et al.* 2013) was enriched, as was retinol metabolism, consistent with cadmium responses observed in *C. elegans* (Cui & Freedman 2009).

KEGG Path	Genes	Enrichment	p-value	Activation	p-value
Pentose and glucuronate interconversions	31	42.77	4.62E-06	0.71	2.94E-02
Glutathione metabolism	81	19.19	1.12E-05	0.60	7.48E-02
Arachidonic acid metabolism	41	40.55	3.68E-07	0.56	5.33E-01
Retinol metabolism	28	34.27	1.39E-04	0.61	3.45E-01
Metabolism of xenobiotics by cytochrome P450	47	43.25	1.77E-08	0.66	4.00E-02
Drug metabolism - cytochrome P450	45	45.40	1.35E-08	0.64	7.25E-02

Table 4-2 Sensitive Response to Cadmium: Pathway enrichment and activation of six KEGG paths that were significantly enriched (FDR < 0.05) in genes that responded to cadmium in the sensitive clone. The number of genes on the path, the enrichment rate, enrichment p value, activation rate, and activation p value is shown. Activation rate specifies the fraction of significant genes that were induced in response to cadmium in the sensitive clone, and the activation p value is based on a binomial test in which the null hypothesis is that the activation rate is 0.5.

KEGG Path	Genes	Enrichment	p-value	Activation	p-value
Pentose and glucuronate interconversions	31	52.94	6.67E-09	0.61	2.81E-01
Ascorbate and aldarate metabolism	24	43.38	4.74E-06	0.67	1.52E-01
Taurine and hypotaurine metabolism	11	47.42	1.19E-03	0.91	1.17E-02
Glutathione metabolism	81	36.26	1.54E-13	0.58	1.82E-01
Arachidonic acid metabolism	41	37.83	3.92E-08	0.61	2.11E-01
Retinol metabolism	28	47.62	1.98E-07	0.61	3.45E-01
Porphyrin and chlorophyll metabolism	32	30.99	1.56E-05	0.44	5.97E-01
Metabolism of xenobiotics by cytochrome P450	47	53.67	1.04E-12	0.60	2.43E-01
Drug metabolism - cytochrome P450	45	56.65	6.79E-13	0.58	3.71E-01
Drug metabolism - other enzymes	40	24.11	3.84E-05	0.55	6.36E-01
Metabolic pathways	1087	5.40	1.39E-09	0.40	1.41E-11
Ribosome	124	9.17	3.16E-04	0.05	4.42E-28

Table 4-3 Tolerant Response to Cadmium: Pathway enrichment and activation of six KEGG paths that were significantly enriched (FDR < 0.05) in genes responded to cadmium in the tolerant clone. The number of genes on the path, the enrichment rate, enrichment p value, activation rate, and activation p value is shown. Activation rate specify the fraction of significant genes that were induced in response to cadmium the tolerant, and the activation p value is based on a binomial test in which the null hypothesis is that the activation rate is 0.5.

Tolerant Clone Activates Taurine and Represses Ribosome Pathway

The tolerant clone regulated more genes significantly in response to cadmium, 117 as opposed to 89, translating into the larger number of enriched KEGG pathways in Table 3. Pathways that were significantly enriched only in the tolerant clone included ascorbate and aldarate metabolism, taurine metabolism, porphyrin and chlorophyll metabolism, "Drug metabolism – other enzymes", the general "metabolic pathways" pathway, and the pathway KEGG refers to simply as "ribosome". The taurine path, which was significantly activated (p = 1.17e-02), and the ribosome path, which was significantly repressed (p = 4.42e-28) may contribute to cadmium tolerance. Taurine blocks metal-induced neurotoxicity (Sinha *et al.* 2008; Xu *et al.* 2015) possibly by reducing oxidative stress in mitochondria. Reduced ribosomal gene expression leads to reduced protein translation, a protective stress response (Yamasaki & Anderson 2008). Reduced protein synthesis has been observed in response to

heavy metal stresses including cadmium (Pytharopoulou *et al.* 2011), observed in cadmium-exposed *Daphnia* Connon:2008gf, Shaw:2007ge}and tied to cadmium tolerance (De Coninck *et al.* 2014).

Tolerant and Sensitive Clones Correlated in Cadmium Expression Response

Gene set analysis uses specific thresholds to categorize genes in a binary fashion: a gene reaches FDR < 0.05 or it fails to, but genes falling on either side of an arbitrary threshold value such as FDR = 0.05 may respond similarly. To understand similarities and differences between the two clones in a more realistic way, I determined the level of correlations between tolerant and sensitive clones in their gene expression response to cadmium, as shown in Figure 4-3.

Figure 4-3 was constructed by selecting the paths that were among the top ten most significantly enriched paths in either clone, and adding the taurine and ribosome paths. The figure demonstrates that the two clones generally regulate the same genes in the same direction, regardless of whether those genes achieve significance in both clones. Gene expression response in the sensitive clone (x axis) is plotted against response in the tolerant clone (y axis) for all genes on each path. Blue regression lines, surrounded by 95% confidence intervals, generally show significant and positive correlation in response. Most pathway genes were not significantly differentially expressed in response to cadmium at the FDR < 0.05 level, as indicated by the grey symbols in Figure 2, and many were significantly differentially expressed only in the tolerant clone (blue symbols). However, the majority of genes that were nominally differentially expressed only in the tolerant clone nonetheless responded in the same direction in the sensitive clone, suggesting that both clones use the same genes, regulated in the same way, but the tolerant clone articulates a more robust response. A similar relationship between cadmium tolerance and gene expression response

was observed in *F. candida*, a soil dwelling hexapod, by Nota, Roelofs, et al (Nota *et al.* 2013). They reported that cadmium tolerance was associated with 1) a larger number of cadmium-responsive genes, 2) larger fold changes in these genes, and 3) more "coordinated" gene regulatory networks, as evidenced by higher levels of Pearson correlation among the cadmium responsive genes that differed between sensitive and tolerant clones.

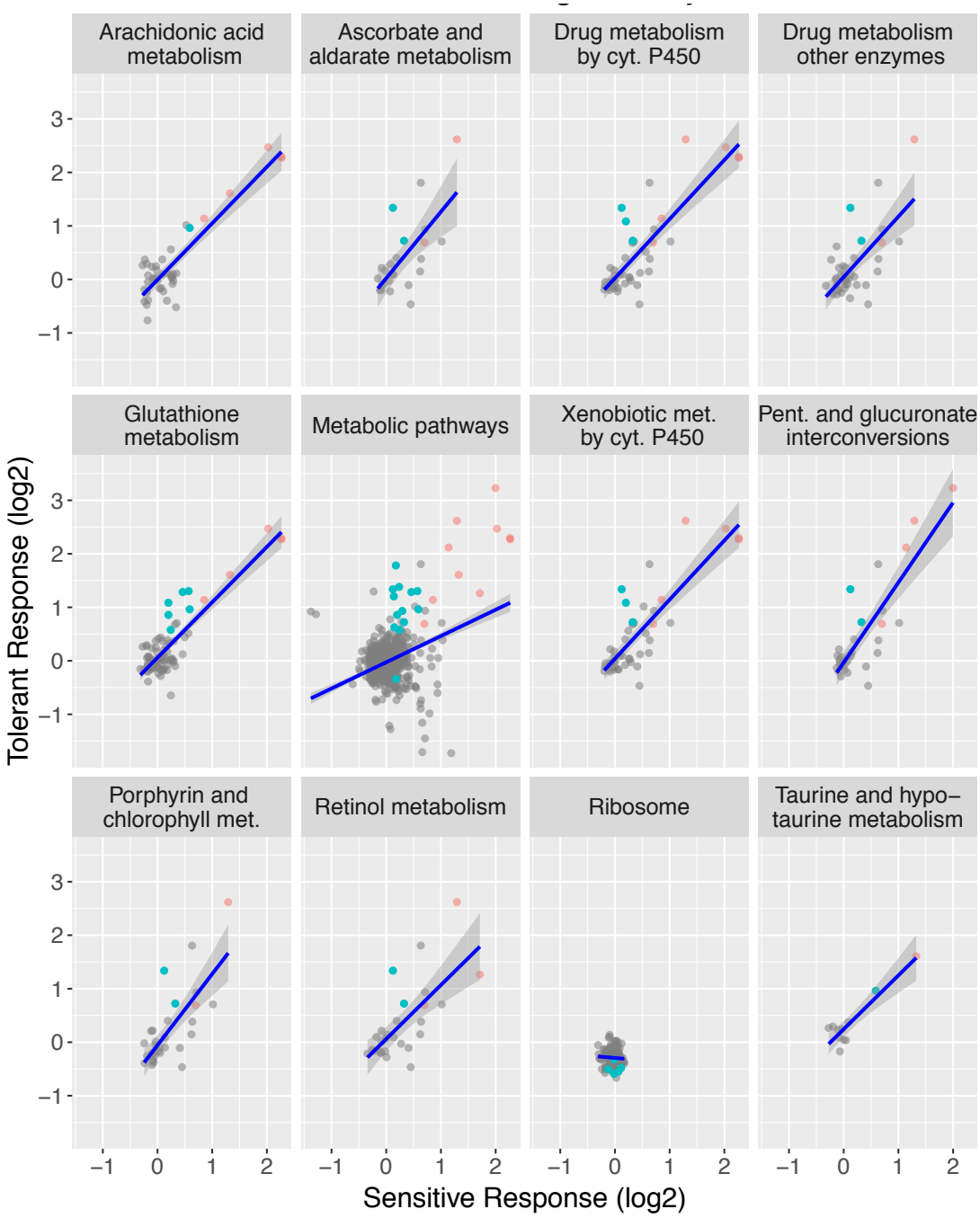

Figure 4-3 Cadmium effects, relative to control, of genes in KEGG pathways that were significantly activated or repressed by cadmium exposure in either the tolerant (y axis) or sensitive (x axis) clone. Grey symbols: pathway genes with FDR > 0.05 in both clones. Red symbols: genes with FDR < 0.05 in both clones. Green symbols: genes with FDR < 0.05 only in the tolerant clone. Blue lines show the linear model fit

regression line surrounded by shaded area that represents the 95% confidence interval of the regression.

Tolerant Response to Cadmium is Systematically Larger

To test the hypothesis that an enhanced gene expression response facilitates cadmium tolerance, I identified three gene sets: the 60 genes that achieved significance (FDR < 0.05) only in the sensitive clone, the 29 genes that achieved significance in both clones, and the 88 genes that achieved significance only in the tolerant clone. As shown in the center panel of Figure 4-4, the responses in the tolerant clone were significantly larger ($p = 0.0005$, paired t test) with an average increase of 31% over the sensitive clone. A modest but significant ($p = 0.03$, Welch's t test) 13% increase was observed in genes that responded only in the tolerant clone (right panel, Figure 4-4) compared to those responding only in the sensitive clone (left panel, Figure 4-4).

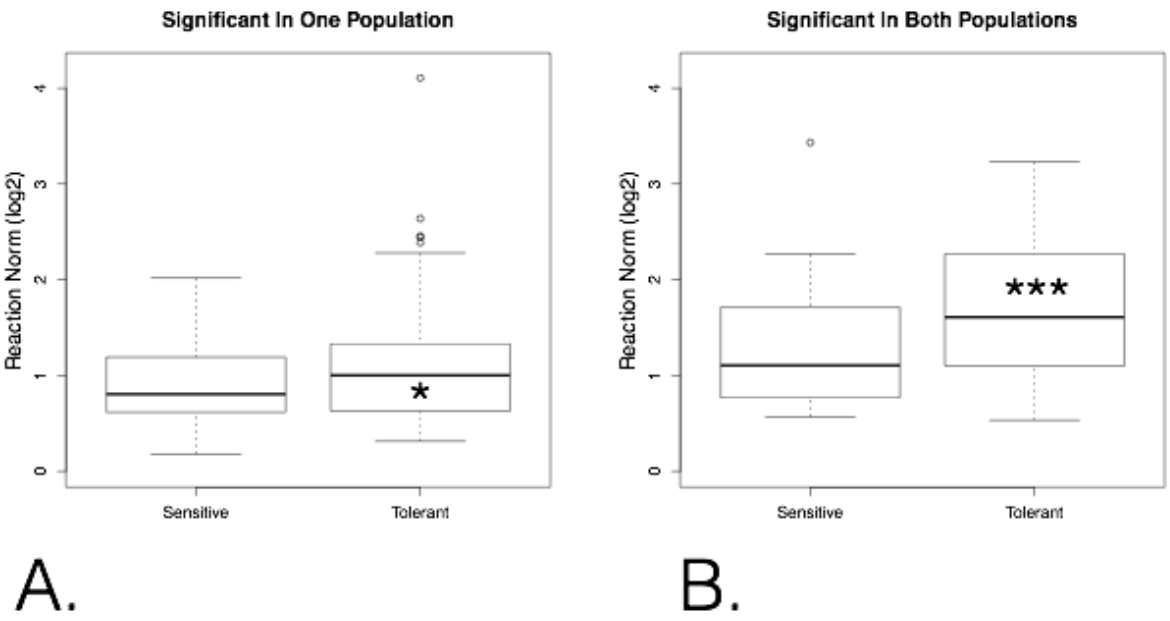

Figure 4-4 Absolute magnitude of cadmium response for all significant genes (FDR < 0.05) in either sensitive or tolerant clones. A: genes with FDR < 0.05 only in one population. B: genes reaching FDR < 0.05 in both clones, *** p = 0.0005 in paired t test. Right: Genes with FDR < 0.05 only in the tolerant clone had significantly larger absolute responses (* p = 0.03, t test) than genes with FDR < 0.05 only in the sensitive clone.

Enhanced Response in Tolerant Clone is Cadmium-specific

If enhanced gene expression response is a hallmark of tolerance, it follows that it should be observed only when the phenotype of tolerance is observed. Therefore, I compared reaction norms in cadmium response to five other single toxicant treatments (Figure 4-5). In each case, the four gene sets used in Figure 4-4 were compared: responses significant only in the sensitive clone (green boxes), responses significant in both the sensitive (pink boxes) and tolerant (red boxes) clones, and those that were significant only in the tolerant clone (blue boxes). The tolerant clone exhibits a more robust response to cadmium than the sensitive clone, consistent with Figure 4-4, but its enhanced response was limited to cadmium, consistent with tolerance phenotype, which does not extend to nickel, copper or arsenic.

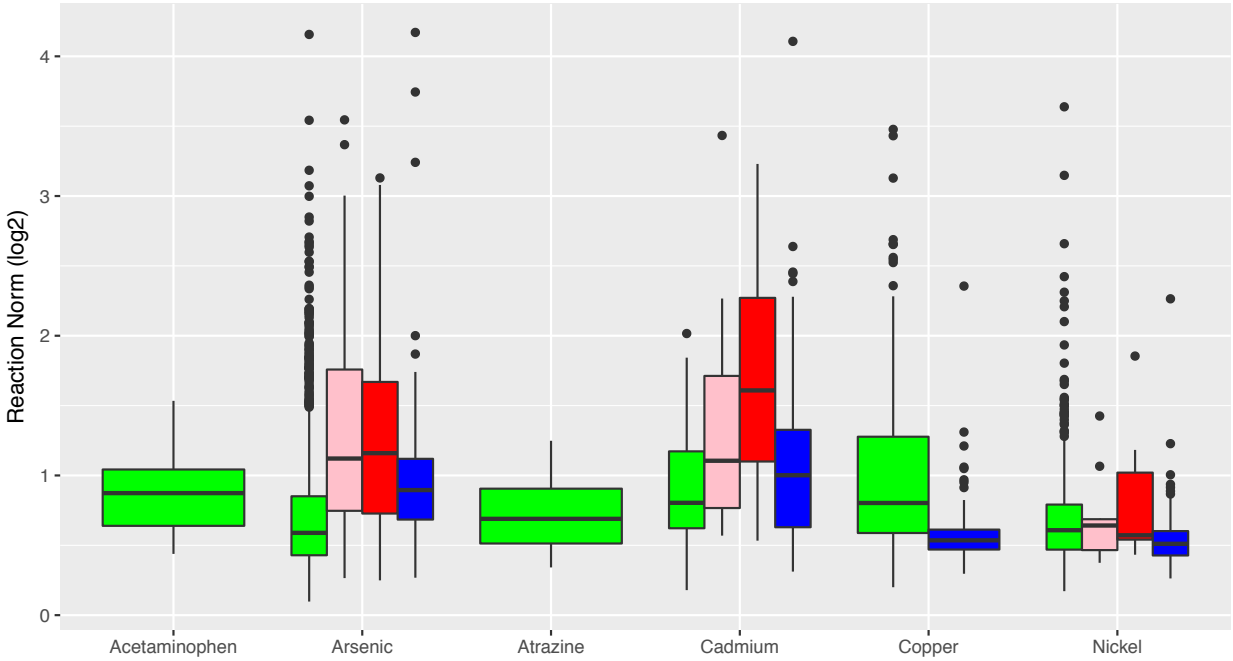

Figure 4-5 Boxplots of absolute difference from control (reaction norm) by toxicant and clone, in $\log_2$ units. Boxes delimit the interquartile range, center lines inside boxes are medians, and whiskers indicate 1.5 times the interquartile range. Green: sensitive clone's response for genes that responded significantly only in the sensitive clone. Blue: tolerant clone's response for genes that responded significantly only in the tolerant clone. Pink: sensitive clone's response for genes that were significantly responsive in both clones. Red: tolerant clone's response for genes that were significantly responsive in both clones.

Co-regulation of cadmium-sensitive genes is higher in tolerant clone

If cadmium tolerance in *D. pulicaria* relies on systematically larger responses to cadmium in cadmium-responsive genes, this might be expected to alter gene expression in networks composed of cadmium responsive genes. I calculated pairwise correlations across all 177 genes that responded significantly (FDR < 0.05) to cadmium, making separate correlation matrices for each clone. The number of significant co-regulation relationships between any

gene and other genes in the same network was independently calculated for each of the 177 genes (see Methods), and the number observed in the sensitive clone were subtracted from the number observed in the tolerant clone, resulting in the distribution shown in Figure 4-6. A minority of genes had fewer significant gene-gene correlations in the tolerant clone than the sensitive one, but on average, as shown by negative values in Figure 4-6. Overall, he tolerant clone averaged 22 additional gene-gene correlations, which was statistically significant (p-value = 2.69e-09, paired t test).

There are many molecular mechanisms that might account for enhanced gene expression responses to cadmium. For example, increased activity of MTF-1 early response genes following cadmium exposure would be predicted to increase transcription of many metal responsive genes such as metallothionein but also glutathione and many others (Janssens *et al.* 2009; Asselman *et al.* 2012; Elran *et al.* 2014). However, MTF-1 was not constitutively over-expressed in the tolerant clone, nor was it more highly expressed in in response to cadmium in the tolerant clone, and finally, increased metallothionein expression was not observed in the tolerant clone compared to the sensitive clone (Supplemental Tables 1, 2). This is consistent with the observation that the tolerant clone is as sensitive to copper as the sensitive clone, because increased activity on the MTF-1 glutathione access would be expected to provide cross tolerance to many metals.

Factors other than MTF-1 signaling, whether they involve other transcription factors, micro RNAs, histone modifications, methylation, or some other mechanism, result in a more coordinated cadmium-responsive networks in the tolerant clone than the sensitive one.

In summary, I found that cadmium tolerance in *D. pulicaria* is probably facilitated by enhanced gene expression responses of genes that participate in highly coordinated gene expression networks. As this network-driven cadmium tolerance has been observed in *F.*

candida, it is likely that further research might identify alterations in network structure as a common adaptive response to environmental stress, generalizing to other species and other toxicants. This type of analytical approach, in which systematic differences in gene regulatory network structure and responsiveness is considered in addition to gene-centric responses, may be useful in understanding and predicting population responses to a variety of current or future environmental stresses.

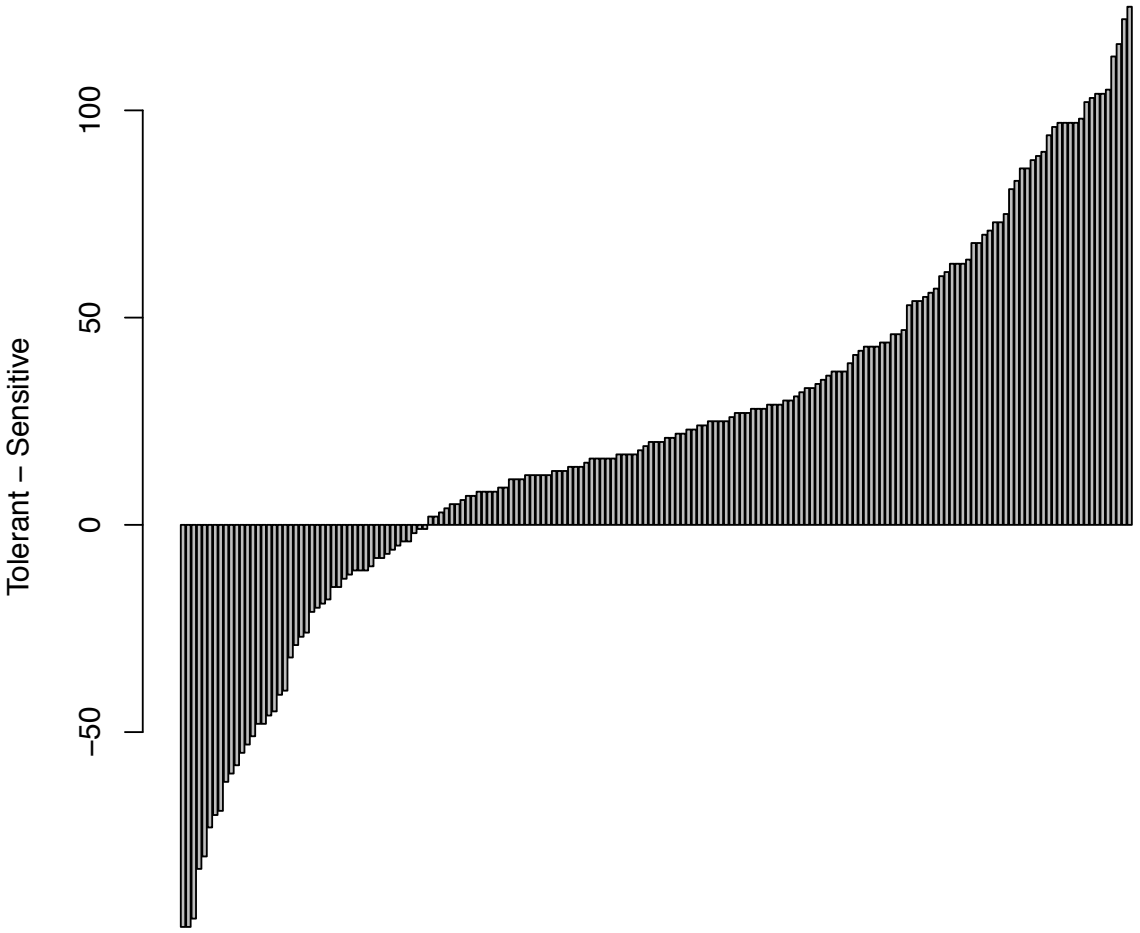

Figure 4-6 Co-regulation among cadmium responsive genes. Clone differences in the number of significant correlations, for each gene responding significantly to cadmium.

Bar heights represent the number of significant correlations in the tolerant clone less the number of significant correlations in the sensitive clone.

Supplemental Section

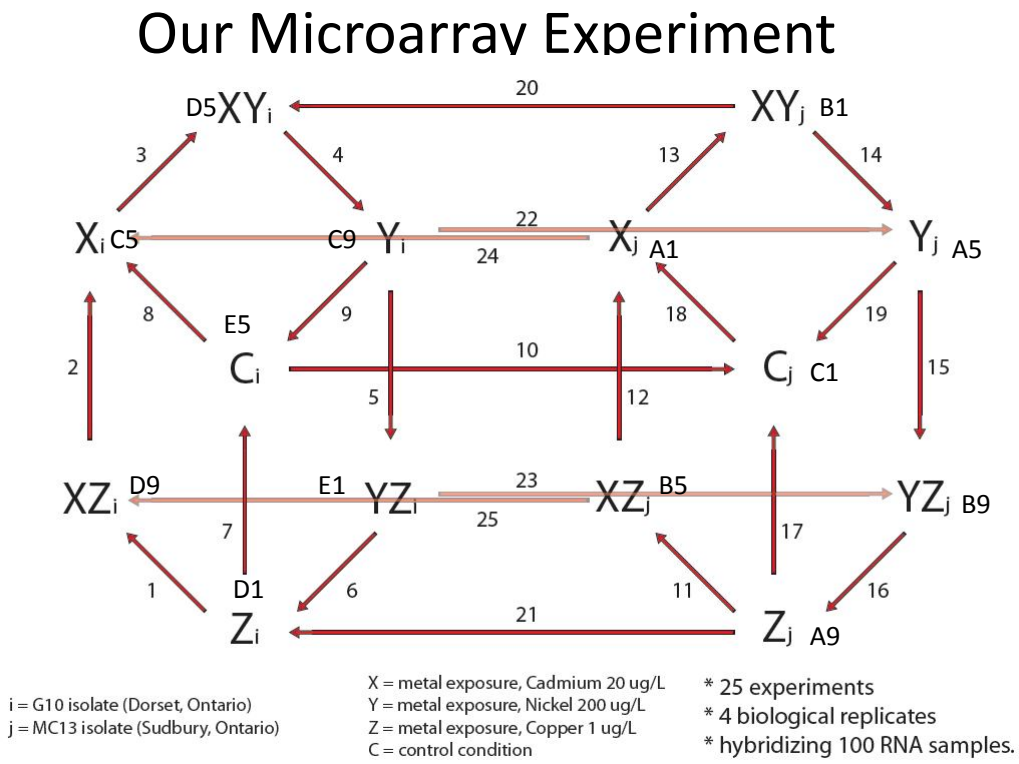

Figure 4-7 (Supplemental) Microarray loop design.

Gene ID	Description	Estimate	P
335886	5-oxoprolinase (ATP-hydrolysing). Glutathione metabolism,Oxoprolinase	-0.506	2.28E-06
304866	5-oxoprolinase (ATP-hydrolysing). Glutathione metabolism,Oxoprolinase	-0.869	4.42E-08
305015	Ascorbate and aldarate metabolism, Cytochrome P450 CYP11/CYP12/CYP24/CYP27 subfamilies	-0.263	1.84E-03
335982	Ascorbate and aldarate metabolism, Cytochrome P450 CYP11/CYP12/CYP24/CYP27 subfamilies	-0.370	1.75E-02
327655	Ascorbate and aldarate metabolism, Cytochrome P450 CYP11/CYP12/CYP24/CYP27 subfamilies	-0.401	1.84E-05
311474	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	2.660	6.29E-10
221284	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	1.393	3.48E-10
311594	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	1.103	4.27E-10
206765	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	0.741	1.14E-03
305686	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	0.680	3.33E-06
311593	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	0.648	5.66E-07
96133	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	0.623	3.97E-03
303449	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	0.511	5.66E-05
50258	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	0.464	3.66E-04
311262	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	-0.290	5.18E-04
235307	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	-0.453	2.78E-04
311387	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	-0.482	4.13E-05
312044	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	-0.599	1.09E-04
311388	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	-0.743	6.13E-08
320218	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	-0.981	7.95E-08
311386	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	-1.199	1.04E-09
311385	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	-1.402	3.56E-10
311767	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	-1.582	1.36E-10
248107	Cu2+/Zn2+ superoxide dismutase SOD1	0.550	2.03E-02
305084	Cu2+/Zn2+ superoxide dismutase SOD1	0.540	3.14E-04
231065	Cu2+/Zn2+ superoxide dismutase SOD1	-0.610	4.20E-07
321908	Cytochrome P450 CYP11/CYP12/CYP24/CYP27 subfamilies	0.692	9.12E-05
309471	Cytochrome P450 CYP2 subfamily	1.657	5.24E-12
301662	Cytochrome P450 CYP2 subfamily	1.438	5.68E-08
315371	Cytochrome P450 CYP2 subfamily	1.365	8.07E-09
194538	Cytochrome P450 CYP2 subfamily	0.911	1.05E-08
234601	Cytochrome P450 CYP2 subfamily	0.788	5.34E-03
300053	Cytochrome P450 CYP2 subfamily	0.448	3.70E-04
314974	Cytochrome P450 CYP2 subfamily	-0.421	8.59E-03
97650	Cytochrome P450 CYP2 subfamily	-0.741	2.62E-07
67675	Cytochrome P450 CYP2 subfamily	-0.971	1.61E-07
259484	Cytochrome P450 CYP3/CYP5/CYP6/CYP9 subfamilies	-0.469	1.35E-03
96262	Cytochrome P450 CYP3/CYP5/CYP6/CYP9 subfamilies	-0.657	6.59E-04
47072	Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	0.660	5.57E-04
42067	Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	-0.476	8.81E-04
192258	Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	-0.794	5.39E-05
53633	Failed axon connections (fax) protein/glutathione S-transferase-like protein	0.991	3.67E-08
303601	Failed axon connections (fax) protein/glutathione S-transferase-like protein	0.329	3.40E-03
246453	Failed axon connections (fax) protein/glutathione S-transferase-like protein	-0.171	2.05E-02
43961	Failed axon connections (fax) protein/glutathione S-transferase-like protein	-0.956	1.35E-08
311604	Fatty acid metabolism, Cytochrome P450 CYP3/CYP5/CYP6/CYP9 subfamilies	1.190	6.84E-12
311606	Fatty acid metabolism, Cytochrome P450 CYP3/CYP5/CYP6/CYP9 subfamilies	0.685	6.87E-07
311603	Fatty acid metabolism, Cytochrome P450 CYP3/CYP5/CYP6/CYP9 subfamilies	0.655	2.83E-06
311602	Fatty acid metabolism, Cytochrome P450 CYP3/CYP5/CYP6/CYP9 subfamilies	-0.440	4.88E-05
225529	Fatty acid metabolism, Cytochrome P450 CYP3/CYP5/CYP6/CYP9 subfamilies	-0.618	5.45E-07
43213	Fatty acid metabolism, P450 CYP2 subfamily	1.449	2.49E-06
310332	Fatty acid metabolism, P450 CYP2 subfamily	0.556	1.63E-05
315215	Fatty acid metabolism, P450 CYP2 subfamily	-0.724	1.25E-06
47261	Fatty acid metabolism, P450 CYP2 subfamily	-1.225	4.80E-13
307855	Fatty acid metabolism, P450 CYP2 subfamily	-1.431	2.18E-10
239125	Fatty acid metabolism, P450 CYP2 subfamily	-2.567	6.02E-16
300931	Gamma-glutamyltransferase. Taurine and hypotaurine metabolism, Glutathione metabolism	-0.524	7.27E-03
112919	Glutathione peroxidase	-0.395	2.05E-04
68047	Glutathione peroxidase. Glutathione metabolism,Glutathione peroxidase	-0.392	5.80E-05
218103	Glutathione peroxidase. Glutathione metabolism,Glutathione peroxidase	-0.588	1.29E-06
114529	Glutathione peroxidase. Glutathione metabolism,Glutathione peroxidase	-0.941	2.78E-07
233749	Glutathione peroxidase. Glutathione metabolism,Glutathione peroxidase	-1.259	1.05E-08
304194	Glutathione peroxidase. Glutathione metabolism,Glutathione S-transferase	-0.362	9.15E-03
196080	Glutathione S-transferase	0.990	3.03E-06
230303	Glutathione S-transferase	-0.197	4.13E-03
318232	Glutathione S-transferase	-0.363	8.53E-03
230650	Glutathione S-transferase	-0.436	5.17E-06
205726	Glutathione S-transferase	-0.545	3.57E-04
255502	Glutathione S-transferase	-0.661	5.87E-03
129499	Glutathione synthase. Glutamate metabolism,Glutathione synthase. Glutathione metabolism,Glutathione synthetase	0.337	5.87E-06
231546	Glutathione transferase. Glutathione metabolism	-0.250	1.21E-02
308406	Glutathione transferase. Glutathione metabolism	-0.521	1.31E-04
316498	Glutathione transferase. Glutathione metabolism,Glutathione S-transferase	1.589	2.43E-09
219658	Glutathione transferase. Glutathione metabolism,Glutathione S-transferase	1.031	2.36E-06
210571	Glutathione transferase. Glutathione metabolism,Glutathione S-transferase	0.488	6.52E-04
123298	Glutathione transferase. Glutathione metabolism,Glutathione S-transferase	-0.283	2.48E-03
223594	Glutathione transferase. Glutathione metabolism,Glutathione S-transferase	-0.308	1.75E-02
241024	Glutathione transferase. Glutathione metabolism,Glutathione S-transferase	-0.357	5.89E-03
305501	Glutathione transferase. Glutathione metabolism,Glutathione S-transferase	-0.383	5.47E-04
234554	Glutathione transferase. Glutathione metabolism,Glutathione S-transferase	-0.391	2.21E-05
317266	Glutathione transferase. Glutathione metabolism,Glutathione S-transferase	-0.403	4.25E-04
318250	Glutathione transferase. Glutathione metabolism,Glutathione S-transferase	-0.837	5.12E-06
303282	Glutathione transferase. Glutathione metabolism,Glutathione S-transferase	-1.071	9.56E-07
324074	Hsp70-interacting protein Hip/Transient component of progesterone receptor complexes and an Hsp70-binding protein	-1.192	2.48E-09
302859	Membrane alanine aminopeptidase. Glutathione metabolism,Puromycin-sensitive aminopeptidase and related aminopeptidases	0.930	1.40E-06
243400	Membrane alanine aminopeptidase. Glutathione metabolism,Puromycin-sensitive aminopeptidase and related aminopeptidases	0.528	3.65E-03
304409	Membrane alanine aminopeptidase. Glutathione metabolism,Puromycin-sensitive aminopeptidase and related aminopeptidases	0.301	3.69E-03
317254	Membrane alanine aminopeptidase. Glutathione metabolism,Puromycin-sensitive aminopeptidase and related aminopeptidases	-0.270	1.36E-02
64107	Membrane alanine aminopeptidase. Glutathione metabolism,Puromycin-sensitive aminopeptidase and related aminopeptidases	-0.923	1.96E-03
331359	Membrane alanine aminopeptidase. Glutathione metabolism,Puromycin-sensitive aminopeptidase and related aminopeptidases	-1.116	6.55E-06
65912	Molecular chaperones HSP105/HSP110/SSE1. HSP70 superfamily	0.364	4.63E-03
302856	Molecular chaperones HSP70/HSC70. HSP70 superfamily	0.266	9.47E-05
96986	Molecular chaperones HSP70/HSC70. HSP70 superfamily	0.239	7.61E-03
271722	Molecular chaperones HSP70/HSC70. HSP70 superfamily	-0.086	5.78E-03
114156	Molecular chaperones HSP70/HSC70. HSP70 superfamily	-0.783	3.65E-08
265325	Molecular chaperones HSP70/HSC70. HSP70 superfamily	-2.314	9.40E-07
227800	Molecular chaperones HSP70/HSC70. HSP70 superfamily	-2.477	1.09E-14
53476	Molecular chaperones HSP70/HSC70. HSP70 superfamily	-4.226	2.98E-19
337893	Multidrug resistance-associated protein/mitoxantrone resistance protein. ABC superfamily	0.848	2.45E-05
308086	Multidrug resistance-associated protein/mitoxantrone resistance protein. ABC superfamily	0.745	6.07E-06
301462	Multidrug resistance-associated protein/mitoxantrone resistance protein. ABC superfamily	0.553	1.22E-02
328283	Multidrug resistance-associated protein/mitoxantrone resistance protein. ABC superfamily	0.308	9.12E-03
330093	Multidrug resistance-associated protein/mitoxantrone resistance protein. ABC superfamily	0.300	6.72E-03
302395	Multidrug resistance-associated protein/mitoxantrone resistance protein. ABC superfamily	-0.535	3.40E-05
18536	Multidrug resistance-associated protein/mitoxantrone resistance protein. ABC superfamily	-0.588	1.25E-04
220769	Phospholipid-hydroperoxide glutathione peroxidase. Glutathione metabolism,Glutathione peroxidase	-0.972	1.34E-08
231071	Predicted metalloprotease with chaperone activity (RNAse H/HSP70 fold)	-0.320	8.12E-03
290507	Predicted Metallothionein Mt2	-0.305	8.91E-03
13439	Predicted Metallothionein Mt4	-0.359	5.22E-03
312222	Predicted MTF-1 transcription factor	-0.083	6.91E-03
316534	Prostaglandin-D synthase. Prostaglandin and leukotriene metabolism,Glutathione S-transferase	-0.273	4.99E-04
305010	Superoxide dismutase. ,Cu2+/Zn2+ superoxide dismutase SOD1	-0.539	6.57E-06
304036	Superoxide dismutase. ,Cu2+/Zn2+ superoxide dismutase SOD1	-0.550	2.48E-03
328913	Thromboxane-A synthase. Prostaglandin and leukotriene metabolism,Cytochrome P450 CYP3/CYP5/CYP6/CYP9 subfamilies	0.165	2.04E-02

Table 4-4 (Supplemental) Estimates of baseline expression differences for genes with annotations matching metallothionein, glutathione, MTF-1, multidrug resistance, cytochrome P450, superoxide dismutase, or heat shock, that differed significantly between sensitive and tolerant clones in baseline expression. *D. pulicaria* gene ID, gene description, $\log_2$ fold change ratio of expression between tolerant and sensitive clones, and p value from linear model comparing baseline gene expression are shown. Strong activation in the tolerant clone ($\log_2$ fold change greater than 0.5) are highlighted in blue, strong repression is highlighted in red

Gene ID	Desc	Sens.E	Sens.P	Tol.E	Tol.P
105886	5-oxoprolinase (ATP-hydrolysing), Glutathione metabolism,Oxoprolinase	-0.162	1.72E-01	0.082	3.46E-01
104866	5-oxoprolinase (ATP-hydrolysing), Glutathione metabolism,Oxoprolinase	0.125	2.70E-01	0.072	5.71E-01
327055	Ascorbate and aldarate metabolism, Cytochrome P450 CYP11/CYP12/CYP24/CYP27 subfamilies	-0.008	9.42E-01	0.073	6.05E-01
305015	Ascorbate and aldarate metabolism, Cytochrome P450 CYP11/CYP12/CYP24/CYP27 subfamilies	0.245	1.09E-02	0.070	6.90E-01
98968	Ascorbate and aldarate metabolism, Cytochrome P450 CYP11/CYP12/CYP24/CYP27 subfamilies	0.064	5.17E-01	0.011	9.21E-01
319971	Ascorbate and aldarate metabolism, Cytochrome P450 CYP11/CYP12/CYP24/CYP27 subfamilies	-0.145	3.58E-01	-0.239	1.14E-01
335982	Ascorbate and aldarate metabolism, Cytochrome P450 CYP11/CYP12/CYP24/CYP27 subfamilies	-0.107	6.08E-01	-0.346	4.37E-02
22745	Ascorbate and aldarate metabolism, Cytochrome P450 CYP2 subfamily	0.772	6.20E-03	-0.152	5.64E-01
306531	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	0.102	6.98E-01	0.732	7.30E-02
320218	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	0.064	6.76E-01	0.788	6.35E-04
309254	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	-0.165	1.77E-01	0.448	2.04E-03
311474	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	-0.040	8.88E-01	0.385	4.15E-01
312044	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	-0.340	2.68E-02	0.281	1.62E-01
221284	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	-0.175	5.06E-01	0.195	3.53E-01
311394	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	-0.259	4.77E-01	0.173	8.08E-01
255307	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	-0.148	4.13E-01	0.169	3.28E-01
320829	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	0.968	8.16E-03	0.159	7.80E-01
96133	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	0.350	2.49E-01	0.158	5.77E-01
311387	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	-0.070	6.39E-01	0.145	3.74E-01
311475	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	0.298	5.06E-02	0.120	7.90E-01
206765	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	0.172	3.70E-01	0.071	8.25E-01
303449	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	0.157	4.23E-01	0.043	7.84E-01
311388	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	-0.118	1.99E-01	-0.016	8.94E-01
311767	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	-0.136	3.76E-01	-0.022	8.95E-01
312048	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	-0.084	5.05E-01	-0.024	8.61E-01
311473	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	0.195	5.30E-01	-0.063	8.53E-01
50258	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	0.095	4.99E-01	-0.086	5.31E-01
311385	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	-0.210	2.66E-01	-0.092	6.81E-01
311383	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	0.021	9.08E-01	-0.098	6.67E-01
311386	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	0.021	8.59E-01	-0.116	4.35E-01
42950	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	-0.056	8.07E-01	-0.119	6.28E-01
311596	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	0.069	8.03E-01	-0.126	7.70E-01
311262	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	-0.215	7.64E-02	-0.157	9.47E-02
305713	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	-0.039	8.20E-01	-0.172	2.86E-01
311593	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	0.017	8.92E-01	-0.197	1.07E-01
311594	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	0.019	8.86E-01	-0.231	1.36E-01
305686	Ascorbate and aldarate metabolism, Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	0.131	5.54E-01	-0.379	2.12E-02
248107	Cu2+/Zn2+ superoxide dismutase SOD1	0.159	6.38E-01	0.104	7.88E-01
92123	Cu2+/Zn2+ superoxide dismutase SOD1	-0.216	3.31E-01	0.051	8.19E-01
305084	Cu2+/Zn2+ superoxide dismutase SOD1	0.276	1.75E-01	-0.024	9.00E-01
221374	Cu2+/Zn2+ superoxide dismutase SOD1	0.274	9.24E-03	-0.030	8.35E-01
231065	Cu2+/Zn2+ superoxide dismutase SOD1	0.048	7.70E-01	-0.115	2.55E-01
329848	Cu2+/Zn2+ superoxide dismutase SOD1	1.214	2.24E-03	-0.217	6.08E-01
17612	Cu2+/Zn2+ superoxide dismutase SOD1	0.893	2.27E-02	-0.220	5.75E-01
229959	Cu2+/Zn2+ superoxide dismutase SOD1	0.352	7.90E-02	-0.384	1.45E-01
226068	Cu2+/Zn2+ superoxide dismutase SOD1	0.342	2.08E-02	-0.532	1.47E-03
321908	Cytochrome P450 CYP11/CYP12/CYP24/CYP27 subfamilies	-0.005	9.66E-01	-0.134	5.49E-01
309471	Cytochrome P450 CYP2 subfamily	-0.038	8.23E-01	1.690	2.33E-07
194538	Cytochrome P450 CYP2 subfamily	0.148	2.58E-01	1.632	5.80E-09
301663	Cytochrome P450 CYP2 subfamily	0.377	6.69E-02	0.910	1.02E-03
314974	Cytochrome P450 CYP2 subfamily	0.056	6.41E-01	0.388	2.49E-01
234601	Cytochrome P450 CYP2 subfamily	0.748	5.38E-02	0.380	3.75E-01
67675	Cytochrome P450 CYP2 subfamily	-0.311	1.94E-01	0.237	2.60E-01
203641	Cytochrome P450 CYP2 subfamily	0.026	8.62E-01	0.159	4.39E-01
315371	Cytochrome P450 CYP2 subfamily	-0.190	1.36E-01	0.049	6.92E-01
97050	Cytochrome P450 CYP2 subfamily	-0.028	8.27E-01	0.034	7.08E-01
301662	Cytochrome P450 CYP2 subfamily	-0.139	3.23E-01	0.009	9.39E-01
300053	Cytochrome P450 CYP2 subfamily	0.387	3.63E-02	-0.007	9.69E-01
96282	Cytochrome P450 CYP3/CYP5/CYP6/CYP9 subfamilies	-0.114	7.02E-01	0.329	2.45E-01
259484	Cytochrome P450 CYP3/CYP5/CYP6/CYP9 subfamilies	-0.079	7.07E-01	0.220	3.05E-01
329890	Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	0.256	3.16E-01	0.309	1.61E-01
42067	Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	-0.218	6.17E-02	0.289	1.14E-01
276730	Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	1.226	3.93E-02	0.286	5.10E-01
47072	Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	0.113	5.96E-01	0.230	2.53E-01
192258	Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	-0.105	5.46E-01	0.186	4.27E-01
42144	Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	-0.050	7.46E-01	0.002	9.92E-01
42254	Cytochrome P450 CYP4/CYP19/CYP26 subfamilies	0.201	3.60E-01	-0.098	7.66E-01
192692	Failed axon connections (fax) protein/glutathione S-transferase-like protein	1.249	7.27E-14	2.886	1.73E-09
43961	Failed axon connections (fax) protein/glutathione S-transferase-like protein	0.107	6.07E-01	2.239	5.16E-07
304539	Failed axon connections (fax) protein/glutathione S-transferase-like protein	-0.159	3.46E-01	0.239	5.35E-02
53633	Failed axon connections (fax) protein/glutathione S-transferase-like protein	0.187	2.97E-01	0.081	6.17E-01
303601	Failed axon connections (fax) protein/glutathione S-transferase-like protein	-0.274	1.11E-02	0.039	7.38E-01
246453	Failed axon connections (fax) protein/glutathione S-transferase-like protein	-0.073	5.02E-01	-0.209	3.94E-04
97346	Fatty acid metabolism, Cytochrome P450 CYP3/CYP5/CYP6/CYP9 subfamilies	0.307	4.28E-02	0.568	9.20E-04
221817	Fatty acid metabolism, Cytochrome P450 CYP3/CYP5/CYP6/CYP9 subfamilies	0.254	8.56E-02	0.502	1.80E-03
221291	Fatty acid metabolism, Cytochrome P450 CYP3/CYP5/CYP6/CYP9 subfamilies	0.122	3.77E-01	0.409	2.68E-02
311606	Fatty acid metabolism, Cytochrome P450 CYP3/CYP5/CYP6/CYP9 subfamilies	-0.036	7.91E-01	0.321	7.51E-03
311602	Fatty acid metabolism, Cytochrome P450 CYP3/CYP5/CYP6/CYP9 subfamilies	-0.176	1.71E-01	0.293	1.75E-02
311604	Fatty acid metabolism, Cytochrome P450 CYP3/CYP5/CYP6/CYP9 subfamilies	-0.152	2.37E-01	0.189	8.95E-02
311603	Fatty acid metabolism, Cytochrome P450 CYP3/CYP5/CYP6/CYP9 subfamilies	-0.121	4.15E-01	0.072	6.95E-01
225529	Fatty acid metabolism, Cytochrome P450 CYP3/CYP5/CYP6/CYP9 subfamilies	-0.036	7.03E-01	0.051	8.06E-01
314153	Fatty acid metabolism, P450 CYP2 subfamily	0.246	1.82E-01	0.269	2.84E-01
47355	Fatty acid metabolism, P450 CYP2 subfamily	-0.159	4.20E-01	0.072	7.85E-01
46010	Fatty acid metabolism, P450 CYP2 subfamily	0.015	9.39E-01	0.060	7.77E-01
47261	Fatty acid metabolism, P450 CYP2 subfamily	-0.142	2.12E-01	-0.013	9.34E-01
239125	Fatty acid metabolism, P450 CYP2 subfamily	0.004	9.83E-01	-0.025	8.68E-01
310332	Fatty acid metabolism, P450 CYP2 subfamily	-0.029	8.41E-01	-0.071	6.06E-01
47245	Fatty acid metabolism, P450 CYP2 subfamily	-0.102	5.90E-01	-0.195	1.48E-01
315215	Fatty acid metabolism, P450 CYP2 subfamily	0.081	5.82E-01	-0.241	1.37E-01
43213	Fatty acid metabolism, P450 CYP2 subfamily	0.472	6.94E-02	-0.297	4.67E-01
307855	Fatty acid metabolism, P450 CYP2 subfamily	0.173	2.31E-01	-0.397	4.59E-02
258510	Gamma-glutamyltransferase, Taurine and hypotaurine metabolism, Glutathione metabolism	-0.196	3.55E-01	0.293	3.46E-01
300931	Gamma-glutamyltransferase, Taurine and hypotaurine metabolism, Glutathione metabolism	-0.274	1.17E-01	0.264	2.01E-01
300820	Glutamate--cysteine ligase, Glutamate metabolism,Glutamate--cysteine ligase, Glutathione metabolism,Gamma-glutamylcysteine synthetase	0.460	6.35E-03	1.289	5.13E-06
112919	Glutathione peroxidase	-0.172	3.54E-01	-0.389	5.97E-03
218103	Glutathione peroxidase, Glutathione metabolism,Glutathione peroxidase	-0.172	1.83E-01	0.075	6.13E-01
114529	Glutathione peroxidase, Glutathione metabolism,Glutathione peroxidase	-0.080	5.81E-01	-0.001	9.91E-01
233749	Glutathione peroxidase, Glutathione metabolism,Glutathione peroxidase	-0.105	5.95E-01	-0.136	3.07E-01
310801	Glutathione peroxidase, Glutathione metabolism,Glutathione peroxidase	0.185	4.45E-01	-0.163	4.95E-01
68047	Glutathione peroxidase, Glutathione metabolism,Glutathione peroxidase	0.016	8.18E-01	-0.239	2.60E-02
304194	Glutathione peroxidase, Glutathione metabolism,Glutathione S-transferase	0.460	2.93E-02	0.122	6.29E-01
230826	Glutathione S-transferase	0.353	9.97E-03	0.795	2.04E-03
196080	Glutathione S-transferase	0.745	1.68E-02	0.512	7.99E-03
95675	Glutathione S-transferase	0.027	8.88E-01	0.319	1.24E-01
255502	Glutathione S-transferase	-0.144	7.82E-01	0.091	8.42E-01
230303	Glutathione S-transferase	-0.017	8.43E-01	-0.011	8.48E-01
205726	Glutathione S-transferase	-0.084	6.42E-01	-0.041	7.84E-01
230761	Glutathione S-transferase	-0.175	2.50E-01	-0.046	7.52E-01
200523	Glutathione S-transferase	-0.100	4.34E-01	-0.112	3.73E-01
318232	Glutathione S-transferase	-0.121	4.58E-01	-0.119	4.59E-01
230050	Glutathione S-transferase	-0.098	3.66E-01	-0.124	2.70E-01
15523	Glutathione S-transferase	-0.104	2.87E-01	-0.295	2.38E-02
129499	Glutathione synthase, Glutamate metabolism,Glutathione synthase, Glutathione metabolism,Glutathione synthetase	0.061	4.76E-01	0.435	1.25E-03
308406	Glutathione transferase, Glutathione metabolism	0.000	9.99E-01	-0.124	3.83E-01
231546	Glutathione transferase, Glutathione metabolism	0.056	6.98E-01	-0.160	1.20E-01
231548	Glutathione transferase, Glutathione metabolism	-0.194	2.74E-01	-0.191	2.42E-01
210571	Glutathione transferase, Glutathione metabolism,Glutathione S-transferase	2.624	5.08E-13	2.470	1.30E-08
219858	Glutathione transferase, Glutathione metabolism,Glutathione S-transferase	2.258	6.30E-12	2.298	1.52E-07
316498	Glutathione transferase, Glutathione metabolism,Glutathione S-transferase	2.180	1.07E-11	2.273	1.70E-06
303282	Glutathione transferase, Glutathione metabolism,Glutathione S-transferase	0.164	3.76E-01	0.999	6.55E-02
223594	Glutathione transferase, Glutathione metabolism,Glutathione S-transferase	0.096	5.44E-01	0.694	6.32E-01
223730	Glutathione transferase, Glutathione metabolism,Glutathione S-transferase	0.263	1.76E-01	0.045	8.52E-01
302333	Glutathione transferase, Glutathione metabolism,Glutathione S-transferase	0.263	6.77E-02	0.026	8.62E-01
277532	Glutathione transferase, Glutathione metabolism,Glutathione S-transferase	-0.002	9.79E-01	-0.021	6.82E-01
241024	Glutathione transferase, Glutathione metabolism,Glutathione S-transferase	0.292	3.07E-02	-0.027	9.02E-01
234554	Glutathione transferase, Glutathione metabolism,Glutathione S-transferase	-0.070	5.35E-01	-0.044	6.72E-01
305501	Glutathione transferase, Glutathione metabolism,Glutathione S-transferase	0.185	9.77E-02	-0.078	5.31E-01
123298	Glutathione transferase, Glutathione metabolism,Glutathione S-transferase	-0.058	6.67E-01	-0.131	3.51E-01
318250	Glutathione transferase, Glutathione metabolism,Glutathione S-transferase	0.025	8.42E-01	-0.139	4.07E-01
317266	Glutathione transferase, Glutathione metabolism,Glutathione S-transferase	0.134	4.90E-01	-0.320	3.24E-02
324074	Hsp70-interacting protein Hip/Transient component of progesterone receptor complexes and an Hsp70-binding protein	-0.338	2.01E-01	0.021	7.67E-01
307492	Hsp70-interacting protein Hip/Transient component of progesterone receptor complexes and an Hsp70-binding protein	0.039	6.99E-01	-0.167	4.19E-02
226928	Hsp70-interacting protein Hip/Transient component of progesterone receptor complexes and an Hsp70-binding protein	-0.142	1.62E-01	-0.188	8.32E-02
311442	Hydroxyacylglutathione hydrolase, Pyruvate metabolism,Glyoxylase	0.024	8.67E-01	-0.052	7.42E-01
299809	Isocitrate dehydrogenase (NADP+), Glutathione metabolism	0.058	6.70E-01	-0.285	6.61E-02
231858	Lactoylglutathione lyase, Pyruvate metabolism,Glyoxylase	0.095	3.30E-01	-0.095	2.12E-01
302859	Membrane alanine aminopeptidase, Glutathione metabolism,Puromycin-sensitive aminopeptidase and related aminopeptidases	0.017	9.19E-01	0.416	4.43E-03
241400	Membrane alanine aminopeptidase, Glutathione metabolism,Puromycin-sensitive aminopeptidase and related aminopeptidases	0.298	3.21E-01	0.368	3.16E-01
304409	Membrane alanine aminopeptidase, Glutathione metabolism,Puromycin-sensitive aminopeptidase and related aminopeptidases	-0.205	2.41E-01	0.273	5.55E-02
317254	Membrane alanine aminopeptidase, Glutathione metabolism,Puromycin-sensitive aminopeptidase and related aminopeptidases	0.181	1.74E-01	0.169	1.49E-01
329565	Membrane alanine aminopeptidase, Glutathione metabolism,Puromycin-sensitive aminopeptidase and related aminopeptidases	0.023	8.90E-01	0.060	7.14E-01
61964	Membrane alanine aminopeptidase, Glutathione metabolism,Puromycin-sensitive aminopeptidase and related aminopeptidases	-0.192	2.02E-01	0.052	6.61E-01
302210	Membrane alanine aminopeptidase, Glutathione metabolism,Puromycin-sensitive aminopeptidase and related aminopeptidases	0.147	2.32E-01	0.037	8.67E-01
309681	Membrane alanine aminopeptidase, Glutathione metabolism,Puromycin-sensitive aminopeptidase and related aminopeptidases	-0.038	8.30E-01	0.029	8.64E-01
115736	Membrane alanine aminopeptidase, Glutathione metabolism,Puromycin-sensitive aminopeptidase and related aminopeptidases	0.293	1.93E-02	0.006	9.82E-01
317978	Membrane alanine aminopeptidase, Glutathione metabolism,Puromycin-sensitive aminopeptidase and related aminopeptidases	0.123	4.29E-01	-0.002	9.87E-01
64107	Membrane alanine aminopeptidase, Glutathione metabolism,Puromycin-sensitive aminopeptidase and related aminopeptidases	0.324	2.08E-01	-0.100	8.51E-01
305611	Membrane alanine aminopeptidase, Glutathione metabolism,Puromycin-sensitive aminopeptidase and related aminopeptidases	0.121	6.42E-01	-0.154	6.22E-01
331359	Membrane alanine aminopeptidase, Glutathione metabolism,Puromycin-sensitive aminopeptidase and related aminopeptidases	0.089	7.08E-01	-0.189	5.33E-01
305735	Membrane alanine aminopeptidase, Glutathione metabolism,Puromycin-sensitive aminopeptidase and related aminopeptidases	0.246	3.86E-01	-0.647	1.59E-01
63170	Metalloendopeptidase, Predicted metalloprotease with chaperone activity (RNAse H/HSP70 fold)	-0.072	7.71E-01	-0.314	2.39E-01
306788	Molecular chaperones GRP170/SIL1. HSP70 superfamily	-0.031	8.74E-01	-0.083	5.54E-01
304735	Molecular chaperones GRP78/BiP/KAR2. HSP70 superfamily	-0.067	5.68E-01	-0.073	4.46E-01
34365	Molecular chaperones HSP105/HSP110/SSE1. HSP70 superfamily	0.011	9.67E-01	0.161	4.24E-01
229110	Molecular chaperones HSP105/HSP110/SSE1. HSP70 superfamily	-0.081	4.99E-01	0.072	5.76E-01
19575	Molecular chaperones HSP105/HSP110/SSE1. HSP70 superfamily	0.027	7.08E-01	-0.012	8.41E-01
65912	Molecular chaperones HSP105/HSP110/SSE1. HSP70 superfamily	-0.053	8.20E-01	-0.062	7.28E-01
322864	Molecular chaperones HSP105/HSP110/SSE1. HSP70 superfamily	0.025	8.80E-01	-0.126	5.81E-01
227800	Molecular chaperones HSP70/HSC70. HSP70 superfamily	-0.048	6.46E-01	0.163	3.65E-01
69734	Molecular chaperones HSP70/HSC70. HSP70 superfamily	-0.093	3.05E-01	0.103	3.55E-01
96986	Molecular chaperones HSP70/HSC70. HSP70 superfamily	0.030	7.90E-01	0.079	5.84E-01
53625	Molecular chaperones HSP70/HSC70. HSP70 superfamily	0.132	5.54E-01	0.078	6.78E-01
265325	Molecular chaperones HSP70/HSC70. HSP70 superfamily	-0.167	2.03E-01	0.064	8.65E-01
53476	Molecular chaperones HSP70/HSC70. HSP70 superfamily	0.227	2.74E-01	0.064	5.16E-01
271722	Molecular chaperones HSP70/HSC70. HSP70 superfamily	-0.040	6.60E-01	0.031	6.88E-01
306599	Molecular chaperones HSP70/HSC70. HSP70 superfamily	0.132	3.86E-01	0.022	8.61E-01
256736	Molecular chaperones HSP70/HSC70. HSP70 superfamily	-0.010	9.02E-01	-0.011	8.89E-01
302856	Molecular chaperones HSP70/HSC70. HSP70 superfamily	0.115	2.87E-01	-0.026	8.10E-01
114156	Molecular chaperones HSP70/HSC70. HSP70 superfamily	-0.060	6.63E-01	-0.035	8.23E-01
313359	Molecular chaperones mortalin/PBP74/GRP75. HSP70 superfamily	-0.009	9.57E-01	0.165	3.39E-01
308086	Multidrug resistance-associated protein/mitoxantrone resistance protein. ABC superfamily	0.745	9.49E-04	1.166	7.34E-08
326014	Multidrug resistance-associated protein/mitoxantrone resistance protein. ABC superfamily	0.284	9.38E-02	0.366	2.83E-02
18538	Multidrug resistance-associated protein/mitoxantrone resistance protein. ABC superfamily	0.090	4.61E-01	0.206	1.53E-01
330093	Multidrug resistance-associated protein/mitoxantrone resistance protein. ABC superfamily	0.207	1.13E-01	0.197	2.34E-01
337893	Multidrug resistance-associated protein/mitoxantrone resistance protein. ABC superfamily	0.109	6.33E-01	0.160	3.07E-01
302395	Multidrug resistance-associated protein/mitoxantrone resistance protein. ABC superfamily	0.104	3.88E-01	0.154	7.02E-02
328283	Multidrug resistance-associated protein/mitoxantrone resistance protein. ABC superfamily	0.126	2.61E-01	0.071	5.62E-01
306075	Multidrug resistance-associated protein/mitoxantrone resistance protein. ABC superfamily	0.007	9.16E-01	0.068	1.82E-01
308146	Multidrug resistance-associated protein/mitoxantrone resistance protein. ABC superfamily	-0.080	3.08E-01	0.049	6.02E-01
301462	Multidrug resistance-associated protein/mitoxantrone resistance protein. ABC superfamily	0.573	9.44E-02	0.024	9.58E-01
19640	Multidrug resistance-associated protein/mitoxantrone resistance protein. ABC superfamily	-0.166	2.89E-02	-0.005	9.58E-01
19767	Multidrug resistance-associated protein/mitoxantrone resistance protein. ABC superfamily	-0.057	6.98E-01	-0.071	3.89E-01
127589	Multidrug/pheromone exporter, ABC superfamily	0.650	9.83E-05	1.320	4.43E-06
313130	Multidrug/pheromone exporter, ABC superfamily	0.104	3.33E-01	0.161	6.74E-02
66764	Multidrug/pheromone exporter, ABC superfamily	-0.057	7.93E-01	-0.281	2.54E-01
220769	Phospholipid-hydroperoxide glutathione peroxidase, Glutathione metabolism,Glutathione peroxidase	-0.178	3.52E-01	-0.100	5.06E-01
231071	Predicted metalloprotease with chaperone activity (RNAse H/HSP70 fold)	-0.030	8.74E-01	-0.294	6.41E-02
290507	Predicted Metallothionein MI2	0.172	4.37E-01	0.019	9.20E-01
290508	Predicted Metallothionein MI2	1.196	2.21E-03	-0.128	7.11E-01
13439	Predicted Metallothionein MI4	-0.123	2.95E-01	0.058	6.13E-01
312222	Predicted MTF-1 transcription factor	-0.066	4.49E-01	0.044	3.65E-01
307787	Prostaglandin-D synthase. Prostaglandin and leukotriene metabolism,Glutathione S-transferase	1.431	8.60E-07	1.169	5.31E-05
316534	Prostaglandin-D synthase. Prostaglandin and leukotriene metabolism,Glutathione S-transferase	-0.016	8.96E-01	0.173	1.80E-01
56335	Prostaglandin-E synthase. Prostaglandin and leukotriene metabolism,Glutathione S-transferase-related protein	-0.193	9.30E-02	-0.246	1.47E-01
265300	Superoxide dismutase _Copper chaperone for superoxide dismutase	0.016	8.07E-01	-0.127	4.81E-02
66374	Superoxide dismutase _Cu2+/Zn2+ superoxide dismutase SOD1	-0.028	6.03E-01	-0.015	8.21E-01
313514	Superoxide dismutase _Cu2+/Zn2+ superoxide dismutase SOD1	-0.074	4.87E-01	-0.043	7.22E-01
304036	Superoxide dismutase _Cu2+/Zn2+ superoxide dismutase SOD1	0.040	8.30E-01	-0.209	1.25E-01
305010	Superoxide dismutase _Cu2+/Zn2+ superoxide dismutase SOD1	-0.018	8.99E-01	-0.229	1.82E-01
67591	Superoxide dismutase _Manganese superoxide dismutase	-0.082	5.00E-01	-0.153	1.27E-01
96715	Thromboxane-A synthase. Prostaglandin and leukotriene metabolism,Cytochrome P450 CYP3/CYP5/CYP6/CYP9 subfamilies	0.282	6.65E-01	0.097	9.03E-01
328911	Thromboxane-A synthase. Prostaglandin and leukotriene metabolism,Cytochrome P450 CYP3/CYP5/CYP6/CYP9 subfamilies	-0.076	4.42E-01	0.028	8.10E-01

Table 4-5 (Supplemental) Estimates of sensitive and tolerant gene expression responses to cadmium relative to control, for all genes with annotations matching metallothionein, glutathione, MTF-1, multidrug resistance, cytochrome P450, superoxide dismutase, or heat shock proteins, *D. pulicaria* gene ID, description, $\log_2$ fold change ratio of expression and p value sensitive clone, followed by $\log_2$ fold change ratio of expression and p value tolerant clone, from linear model estimates. Strong activation in the tolerant clone ($\log_2$ fold change greater than 0.5) are highlighted in blue, strong repression is highlighted in red.

5. Synopsis

The preceding chapters identified relationships between selective pressures, network cohesion, reaction norms, gene canalization and effective responses to stress. In each case, these relationships were considered in the context of a specific scientific question, such as "how does cadmium tolerance in *D. pulicaria* arise?" or "what is the effect of arsenic exposure on gene expression during acclimation to fresh water in *F. heteroclitus*?" However, since multiple populations of two rather different organisms were studied, it is also possible to interpret these relationships in a more general context and ask "how do these relationships inform our general understanding of evolved stress responses?" An integrated, general model representing the answer to this question is shown in Figure 5-1. This synopsis provides a summary of how results contribute to the general model, identify predictions based on the model, suggest experiments that might test these predictions, and areas of research that might benefit from the methods developed in this thesis.

How Results Contribute to the General Model

Selective Pressures Shape Gene Expression Response Networks

Thousands of genes are differentially expressed between killifish and *Daphnia* populations, yet similar functional pathways respond to osmotic shock in killifish and cadmium response in *Daphnia.* This suggests that selective processes constrain gene expression responses to match functional needs. Selective processes also seem to favor networks with low extramural cohesion (i.e. networks with few regulatory relationships with other networks) in killifish populations responding to hypo-osmotic shock. This validates a previous report (Shaw *et al.* 2014). Finally, levels of gene canalization, i.e., tight gene regulation, were conserved

between populations. A proposed general relationship connecting selective pressures to effective osmotic shock response is shown in Figure 5-1, in blue.

Cadmium Tolerance is Associated with Coordinated, Intense Gene Expression Responses

Cadmium tolerance in *D. pulicaria* is associated with greater coordination of response (increased intramural cohesion) as well as more intense gene expression responses (reaction norms). Similar patterns of increased intramural cohesion and gene expression reaction norm have been previously reported in cadmium tolerant springtails (Nota *et al.* 2013). Collectively, these results are consistent with the view that cadmium exposure exerts selective pressures that favor gene regulatory networks with increased intramural cohesion, facilitating larger responses to cadmium stress, yielding a more effective response. These relationships are shown in Figure 5-1, in green.

Arsenic Degrades Gene Expression Responses and Networks in Killifish

Arsenic disrupts salinity-responsive networks, reducing intramural cohesion. This reduction is coupled with reduced gene expression reaction norms, and both effects are observed at 1 h and 24 h, during both hyper- and hypo-osmotic shock. The connection between intramural cohesion and gene expression reaction norms was also observed in the context of cadmium tolerance in *Daphnia* and springtails adapted to cadmium, consistent with the hypothesis that the two share a causal relationship: perhaps more coordinated gene regulatory networks generally lead to larger reaction norms. Collectively, arsenic interference with gene regulatory networks may explain its ability to reduce the magnitude of gene expression response during osmotic shock, leading to a reduction in the effectiveness of salinity acclimation, as shown by red lines in Figure 5-1.

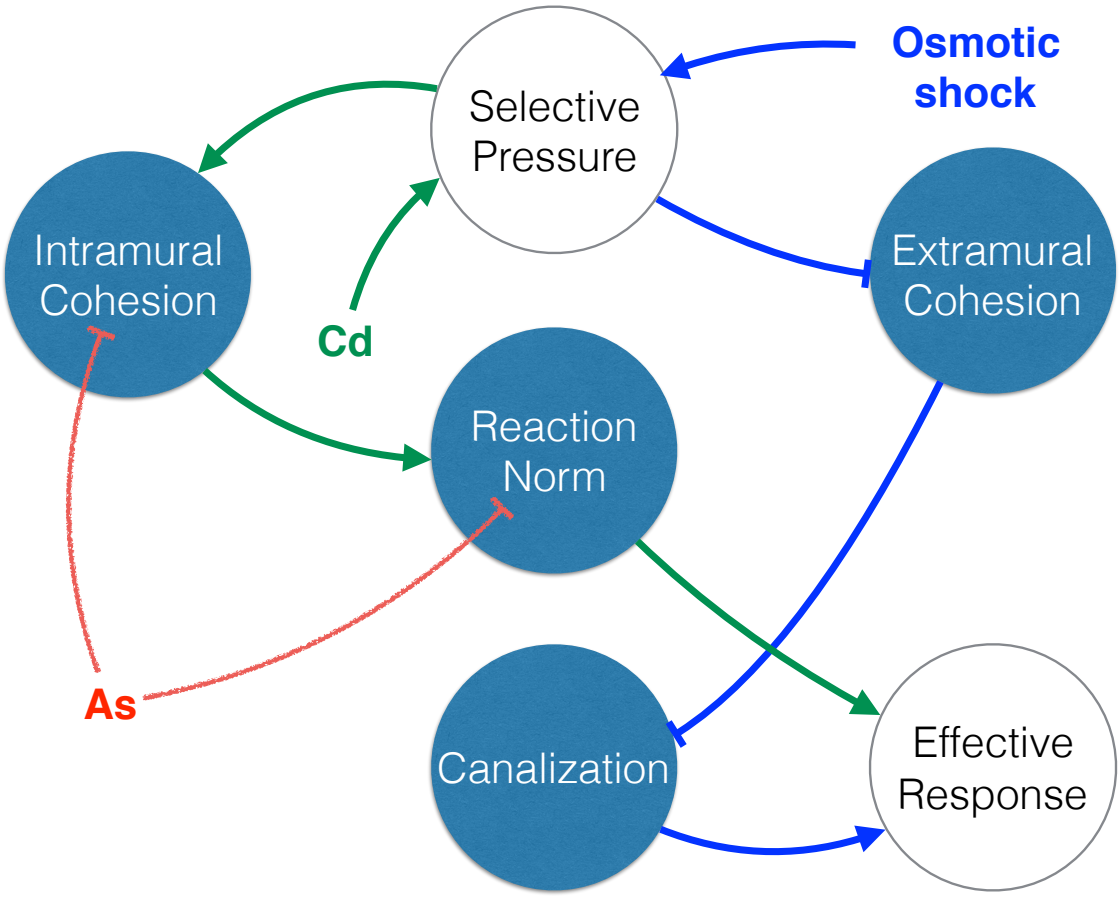

Figure 5-1 A general model outlining relationships between selective pressures, network cohesion, gene expression reaction norms, gene expression canalization, and effective stress responses.

Predictions Based on the General Model

The model presented in Figure 5-1 predicts that other stress-adapted populations should show, in the context of the given response,

- larger gene expression reaction norms,
- networks with higher levels of intramural network cohesion,
- networks with lower levels of extramural network cohesion,
- increased pathway activation, and
- increased gene expression canalization in stress-responsive genes.

In addition, the general model predicts that toxicants other than arsenic might interfere with network structure, and that exposure to arsenic interferes with a variety of stress responses.

Experiments to Test Model Predictions

Some of the experiments that might test the model shown in Figure 5-1 have already been performed. Many cadmium tolerant *Daphnia* populations have been compared to cadmium sensitive *Daphnia* populations since the data presented here were collected, offering an excellent opportunity to observe whether cadmium tolerance in *Daphnia* is facilitated by conserved gene regulatory network structures. In a more general context, cadmium tolerance has been well studied in plants (DalCorso *et al.* 2008), and a meta-analysis of these studies could demonstrate that animals and plants use similar network strategies, such as high intramural cohesion, to adapt to cadmium.

Experiments to validate predicted decreases in intramural cohesion and reaction norms in the presence of arsenic have also been recently performed using additional killifish populations. Generalizing arsenic's ability to impair intramural cohesion and reaction norms could be accomplished by re-analyzing recently published data in which human bronchial epithelial cells were exposed to arsenic during response to the opportunistic pathogen, *Pseudomonas aeruginosa* (Goodale *et al.* 2017).

Applying the Methods Developed in This Thesis to Other Research Domains

Even without verification by additional experiments, this project has already demonstrated the utility of many metrics that capture systematic differences in gene expression responses. These metrics include:

- gene expression reaction norms,
- network cohesion,
- measures of gene expression canalization, and
- gene set activation analysis.

These novel measures of systematic differences in stress responses could improve our understanding of complex problems such as

- the etiology of chronic disease,
- cancer,
- dementia,
- metabolic syndrome, and
- prediction of toxic interactions on the environment.

Conclusions

The overarching question of this thesis project was:

> *How do differences in the magnitude, structure, and coordination of gene expression responses to biotic and abiotic stresses observed in natural populations inform our understanding of adaptive responses, such as acclimation to osmotic shock and cadmium tolerance, and toxic responses, such as the interference of arsenic with*

acclimation to osmotic shock?

This thesis found that:

Coordinated responses to stress with larger magnitudes are associated with cadmium tolerance in Daphnia and more effective acclimation to salinity change in killifish. Tightly controlled genes, organized into networks that share fewer connections with other networks, facilitate osmotic shock responses in killifish. Results gathered in vertebrate and invertebrate wild populations therefore support the views that gene regulatory networks 1) shape the efficacy of biotic and abiotic stress responses, 2) are targeted by toxic effects, and 3) are preserved by selective forces.

References

Allen TEH, Goodman JM, Gutsell S, Russell PJ (2016) A History of the Molecular Initiating Event. *Chemical research in Toxicology*, **29**, 2060–2070.

Andrew AS, Bernardo V, Warnke LA *et al.* (2007) Exposure to Arsenic at Levels Found in U.S. Drinking Water Modifies Expression in the Mouse Lung. *Toxicological Sciences*, **100**, 75–87.

Anglani R, Creanza TM, Liuzzi VC *et al.* (2014) Loss of connectivity in cancer co-expression networks. (F Castiglione, Ed,). *PloS one*, **9**, e87075.

Antczak P, Jo HJ, Woo S *et al.* (2013) Molecular Toxicity Identification Evaluation (mTIE) Approach Predicts Chemical Exposure in *Daphnia magna*. *Environmental Science & Technology*, **47**, 11747–11756.

Apodaca G (2002) Modulation of membrane traffic by mechanical stimuli. *American journal of physiology. Renal physiology*, **282**, F179–F190.

Argos M, Ahsan H, Graziano JH (2012) Arsenic and human health: epidemiologic progress and public health implications. *Reviews on Environmental Health*, **27**, 191–195.

Asselman J, Glaholt SP, Smith Z *et al.* (2012) Functional characterization of four metallothionein genes in *Daphnia pulex* exposed to environmental stressors. *Aquatic Toxicology*, **110-111**, 54–65.

Asselman J, Shaw JR, Glaholt SP, Colbourne JK, De Schamphelaere KAC (2013) Transcription patterns of genes encoding four metallothionein homologs in *Daphnia pulex* exposed to copper and cadmium are time- and homolog-dependent. *Aquatic toxicology (Amsterdam, Netherlands)*, **142-143**, 422–430.

Ayroles JF, Carbone MA, Stone EA *et al.* (2009) Systems genetics of complex traits in

Drosophila melanogaster. *Nature Genetics*, **41**, 299–307.
Bacanskas LR, Whitaker J, Di Giulio RT (2004) Oxidative stress in two populations of killifish (*Fundulus heteroclitus*) with differing contaminant exposure histories. *Marine environmental research*, **58**, 597–601.
Bahrami S, Drabløs F (2016) Gene regulation in the immediate-early response process. *Advances in biological regulation*, **62**, 37–49.
Bain LJ, Liu JT, League RE (2016) Arsenic inhibits stem cell differentiation by altering the interplay between the Wnt3a and Notch signaling pathways. *Toxicology reports*, **3**, 405–413.
Bansal M, Belcastro V, Ambesi-Impiombato A, di Bernardo D (2007) How to infer gene networks from expression profiles. *Molecular Systems Biology*, **3**, 78.
Barata C, Baird DJ, Mitchell SE, Soares AMVM (2002) Among- and within-population variability in tolerance to cadmium stress in natural populations of *Daphnia magna*: implications for ecological risk assessment. *Environmental toxicology and chemistry / SETAC*, **21**, 1058–1064.
Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society Series B (* ….
Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *Annals of statistics*.
Benton MA, Rager JE, Smeester L, Fry RC (2011) Comparative genomic analyses identify common molecular pathways modulated upon exposure to low doses of arsenic and cadmium. *BMC Genomics*, **12**, 173.
Bertin G, Averbeck D (2006) Cadmium: cellular effects, modifications of biomolecules, modulation of DNA repair and genotoxic consequences (a review). *Biochimie*, **88**, 1549–1559.
Bhattacharya S, Bhattacharya A, Roy S (2007) Arsenic-induced responses in freshwater teleosts. *Fish Physiology and Biochemistry*, **33**, 463–473.
Bodar CWM, van der Sluis I, van Montfort JCP, Voogt PA, Zandee DI (1990) Cadmium resistance in *Daphnia magna*. *Aquatic Toxicology*, **16**, 33–39.
Bolstad BM, Bolstad M (2013) Package "preprocessCore."
Bowell RJ, Alpers CN, Jamieson HE, Nordstrom DK, Majzlan J (2014) The Environmental Geochemistry of Arsenic — An Overview —. *Reviews in Mineralogy and Geochemistry*, **79**, 1–16.
Boyce M, Yuan J (2006) Cellular response to endoplasmic reticulum stress: a matter of life or death. *Cell Death & Differentiation*, **13**, 363–373.
Brennan RS, Galvez F, Whitehead A (2015) Reciprocal osmotic challenges reveal mechanisms of divergence in phenotypic plasticity in the killifish *Fundulus heteroclitus*. *The Journal of experimental biology*, jeb.110445.
Bryk J, Somel M, Lorenc A, Teschke M (2013) Early gene expression divergence between allopatric populations of the house mouse (*Mus musculus domesticus*). *Ecology and evolution*, **3**, 558–568.
Bucheli TD, Fent K (2009) Induction of cytochrome P450 as a biomarker for environmental contamination in aquatic ecosystems. *Critical Reviews in Environmental Science and Technology*, **25**, 201–268.
Carter SL, Brechbühler CM, Griffin M, Bond AT (2004) Gene co-expression network topology provides a framework for molecular characterization of cellular state. *Bioinformatics*, **20**, 2242–2250.
Charkoudian N (2003) Skin Blood Flow in Adult Human Thermoregulation: How It Works, When It Does Not, and Why. *Mayo Clinic Proceedings*, **78**, 603–612.

Chen JJ, Hsueh H-M, Delongchamp RR, Lin C-J, Tsai C-A (2007) Reproducibility of microarray data: a further analysis of microarray quality control (MAQC) data. **8**, 412–14.
Chen S, McKinney GJ, Nichols KM, Colbourne JK, Sepúlveda MS (2015) Novel Cadmium Responsive MicroRNAs in *Daphnia pulex*. *Environmental Science & Technology*, **49**, 14605–14613.
Chen S, Nichols KM, Poynton HC, Sepúlveda MS (2016a) MicroRNAs are involved in cadmium tolerance in *Daphnia pulex*. *Aquatic Toxicology*, **175**, 241–248.
Chen Y, Chen C, Tan Z *et al.* (2016b) Functional Identification and Characterization of Genes Cloned from Halophyte Seashore Paspalum Conferring Salinity and Cadmium Tolerance. *Frontiers in Plant Science*, **7**, 306.
Cheng HY, Li P, David M *et al.* (2004) Arsenic inhibition of the JAK-STAT pathway. *Oncogene*, **23**, 3603–3612.
Chiang HS, Guo HR, Hong CL, Lin SM, Lee EF (1993) The incidence of bladder cancer in the black foot disease endemic area in Taiwan. *British journal of urology*, **71**, 274–278.
Ciccarelli RB, Hampton TH, Jennette KW (1981) Nickel carbonate induces DNA-protein crosslinks and DNA strand breaks in rat kidney. *Cancer Letters*, **12**, 349–354.
Clifford M, McGeer JC (2010) Development of a biotic ligand model to predict the acute toxicity of cadmium to *Daphnia pulex*. *Aquatic Toxicology*, **98**, 1–7.
Colbourne JK, Pfrender ME, Gilbert D *et al.* (2011) The Ecoresponsive Genome of Daphnia pulex. *Science*, **331**, 555–561.
Collado-Torres L, Nellore A, Kammers K *et al.* (2017) Reproducible RNA-seq analysis using recount2. *Nature Biotechnology*, **35**, 319–321.
Connon R, Hooper HL, Sibly RM *et al.* (2008) Linking Molecular and Population Stress Responses in *Daphnia magna* exposed to cadmium. *Environmental Science & Technology*.
Consortium SM-I (2014) A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. *Nature Biotechnology*, **32**, 903–914.
Costa D, Mariën J, Janssens TKS *et al.* (2012) Influence of adaptive evolution of cadmium tolerance on neutral and functional genetic variation in *Orchesella cincta*. *Ecotoxicology (London, England)*, **21**, 2078–2087.
Cousyn C, De Meester L, Colbourne JK *et al.* (2001) Rapid, local adaptation of zooplankton behavior to changes in predation pressure in the absence of neutral genetic changes. *Proceedings of the National Academy of Sciences*, **98**, 6256–6260.
Cozzi RRF, Robertson GN, Spieker M *et al.* (2015) Paracellular pathway remodeling enhances sodium secretion by teleost fish in hypersaline environments. *The Journal of experimental biology*, **218**, 1259–1269.
Crawford DL, Oleksiak MF (2007) The biological importance of measuring individual variation. *Journal of Experimental Biology*, **210**, 1613–1621.
Croteau MN, Luoma SN, Stewart AR (2005) Trophic transfer of metals along freshwater food webs: Evidence of cadmium biomagnification in nature. *Limnology and Oceanography*, **50**, 1511–1519.
Crouau Y, Chenon P, Gisclard C (1999) The use of *Folsomia candida* (Collembola, Isotomidae) for the bioassay of xenobiotic substances and soil pollutants. *Applied Soil Ecology*, **12**, 103–111.
Cui Y, Freedman JH (2009) Cadmium Induces Retinoic Acid Signaling by Regulating Retinoic Acid Metabolic Gene Expression. *Journal of Biological Chemistry*, **284**, 24925–24932.
D'haeseleer P, Liang S, Somogyi R (2000) Genetic network inference: from co-expression

clustering to reverse engineering. *Bioinformatics*, **16**, 707–726.
DalCorso G, Farinati S, Maistri S, Furini A (2008) How Plants Cope with Cadmium: Staking All on Metabolism and Gene Expression. *Journal of Integrative Plant Biology*, **50**, 1268–1280.
Davies J, Davies D (2010) Origins and evolution of antibiotic resistance. *Microbiology and molecular biology reviews : MMBR*, **74**, 417–433.
De Coninck DIM, Asselman J, Glaholt S *et al.* (2014) Genome-wide transcription profiles reveal genotype-dependent responses of biological pathways and gene-families in *Daphnia* exposed to single and mixed stressors. *Environmental Science & Technology*, **48**, 3513–3522.
De Coninck DIM, Janssen CR, De Schamphelaere KAC (2013) An investigation of the inter-clonal variation of the interactive effects of cadmium and *Microcystis aeruginosa* on the reproductive performance of *Daphnia magna*. *Aquatic Toxicology*, **140-141**, 425–431.
De Coster S, van Larebeke N (2012) Endocrine-Disrupting Chemicals: Associated Disorders and Mechanisms of Action. *Journal of Environmental and Public Health*, **2012**, 1–52.
de Nadal E, Posas F (2015) Osmostress-induced gene expression--a model to understand how stress-activated protein kinases (SAPKs) regulate transcription. *The FEBS Journal*, **282**, 3275–3285.
de Nadal E, Ammerer G, Posas F (2011) Controlling gene expression in response to stress. *Nature Reviews Genetics*, **12**, 833–845.
Delalande O, Desvaux H, Godat E *et al.* (2010) Cadmium – glutathione solution structures provide new insights into heavy metal detoxification. *The FEBS Journal*, **277**, 5086–5096.
Duker AA, Carranza EJM, Hale M (2005) Arsenic geochemistry and health. *Environment International*, **31**, 631–641.
Duvernell DD, Lindmeier JB, Faust KE, Whitehead A (2008) Relative influences of historical and contemporary forces shaping the distribution of genetic variation in the Atlantic killifish, *Fundulus heteroclitus*. *Molecular Ecology*, **17**, 1344–1360.
Elran R, Raam M, Kraus R *et al.* (2014) Early and late response of *Nematostella vectensis* transcriptome to heavy metals. *Molecular Ecology*, **23**, 4722–4736.
Enard W, Khaitovich P, Klose J *et al.* (2002) Intra- and Interspecific Variation in Primate Gene Expression Patterns. *Science*, **296**, 340–343.
Evans DH (2008) Teleost fish osmoregulation: what have we learned since August Krogh, Homer Smith, and Ancel Keys. *AJP: Regulatory, Integrative and Comparative Physiology*, **295**, R704–R713.
Evans DH (2010) A brief history of the study of fish osmoregulation: the central role of the Mt. Desert Island Biological Laboratory. *Frontiers in physiology*, **1**, 13.
Fangue NA, Hofmeister M, Schulte PM (2006) Intraspecific variation in thermal tolerance and heat shock protein gene expression in common killifish, *Fundulus heteroclitus*. *Journal of Experimental Biology*, **209**, 2859–2872.
Farzan SF, Brickley EB, Li Z *et al.* (2017) Maternal and infant inflammatory markers in relation to prenatal arsenic exposure in a U.S. pregnancy cohort. *Environmental Research*, **156**, 426–433.
Farzan SF, Li Z, Korrick SA *et al.* (2016) Infant Infections and Respiratory Symptoms in Relation to in Utero Arsenic Exposure in a U.S. Cohort. *Environmental Health Perspectives*, **124**, 840–847.
Fiol DF, Kültz D (2007) Osmotic stress sensing and signaling in fishes. *The FEBS Journal*, **274**, 5790–5798.
Fiol DF, Chan SY, Kültz D (2006a) Regulation of osmotic stress transcription factor 1 (Ostf1) in tilapia (*Oreochromis mossambicus*) gill epithelium during salinity stress.

Journal of Experimental Biology, **209**, 3257–3265.
Fiol DF, Chan SY, Kültz D (2006b) Identification and pathway analysis of immediate hyperosmotic stress responsive molecular mechanisms in tilapia (*Oreochromis mossambicus*) gill. *Comparative biochemistry and physiology. Part D, Genomics & proteomics*, **1**, 344–356.
Flint J, Mackay TFC (2009) Genetic architecture of quantitative traits in mice, flies, and humans. *Genome Research*, **19**, 723–733.
Ghosh P, Banerjee M, De Chaudhuri S *et al.* (2007) Comparison of health effects between individuals with and without skin lesions in the population exposed to arsenic through drinking water in West Bengal, India. *Journal of Exposure Science and Environmental Epidemiology*, **17**, 215–223.
Godt J, Scheidig F, Grosse-Siestrup C *et al.* (2006) The toxicity of cadmium and resulting hazards for human health. *Journal of Occupational Medicine and Toxicology*, **1**, 22.
Goodale BC, Rayack EJ, Stanton BA (2017) Arsenic alters transcriptional responses to Pseudomonas aeruginosa infection and decreases antimicrobial defense of human airway epithelial cells. *Toxicology and Applied Pharmacology*.
Goslee SC, Urban DL (2007) The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*.
Gosse JA, Hampton TH, Davey JC, Hamilton JW (2008) *A New Approach to Analysis and Interpretation of Toxicogenomic Gene Expression Data and its Importance in Examining Biological Responses to Low, Environmentally Relevant Doses of Toxicants*. John Wiley & Sons, Ltd, Chichester, UK.
Gregory R, Warnes BB, Lodewijk B (2016) *gplots: Various R programming tools for plotting data*. R package version.
Guan Y, Zhang G-X, Zhang S, Domangue B, Galvez F (2016) The potential role of polyamines in gill epithelial remodeling during extreme hypoosmotic challenges in the Gulf killifish, *Fundulus grandis*. *Comparative biochemistry and physiology. Part B, Biochemistry & molecular biology*.
Haap T, Köhler H-R (2009) Cadmium tolerance in seven *Daphnia magna* clones is associated with reduced hsp70 baseline levels and induction. *Aquatic toxicology (Amsterdam, Netherlands)*, **94**, 131–137.
Haap T, Schwarz S, Köhler H-R (2016) Metallothionein and Hsp70 trade-off against one another in *Daphnia magna* cross-tolerance to cadmium and heat stress. *Aquatic toxicology (Amsterdam, Netherlands)*, **170**, 112–119.
Hampton TH, Stanton BA (2010) A novel approach to analyze gene expression data demonstrates that the ΔF508 mutation in CFTR downregulates the antigen presentation pathway. *AJP: Lung Cellular and Molecular Physiology*, **298**, L473–L482.
Hampton TH, Ballok AE, Bomberger JM *et al.* (2012) Does the F508-CFTR mutation induce a proinflammatory response in human airway epithelial cells? *American journal of physiology. Lung cellular and molecular physiology*, **303**, L509–18.
He S, Jia G, Zhu Z, Tennant DA (2016) Cooperative Co-Evolutionary Module Identification With Application to Cancer Disease Module Discovery - IEEE Xplore Document. *IEEE Transactions In Evolutionary Computation*
Heckel von K, Stephan W, Hutter S (2016) Canalization of gene expression is a major signature of regulatory cold adaptation in temperate *Drosophila melanogaster*. *BMC Genomics*, **17**, 574.
Hedges SB, Marin J, Suleski M, Paymer M, Kumar S (2015) Tree of Life Reveals Clock-Like Speciation and Diversification. *Molecular Biology and Evolution*, msv037.
Hughes MF (2002) Arsenic toxicity and potential mechanisms of action. *Toxicology Letters*, **133**, 1–16.

Hughes MF, Beck BD, Chen Y, Lewis AS, Thomas DJ (2011) Arsenic Exposure and Toxicology: A Historical Perspective. *Toxicological Sciences*, **123**, 305–332.
Hutchinson TC, Whitby LM (1974) Heavy-metal pollution in the Sudbury mining and smelting region of Canada, I. Soil and vegetation contamination by nickel, copper, and other metals. *Environmental Conservation*.
Hwang P-P, Lee T-H (2007) New insights into fish ion regulation and mitochondrion-rich cells. **148**, 479–497.
Isani G, Carpenè E (2014) Metallothioneins, Unconventional Proteins from Unconventional Animals: A Long Journey from Nematodes to Mammals. *Biomolecules*, **4**, 435–457.
Jansen M, Coors A, Stoks R, De Meester L (2011) Evolutionary ecotoxicology of pesticide resistance: a case study in Daphnia. *Ecotoxicology (London, England)*, **20**, 543–551.
Janssens TKS, Roelofs D, van Straalen NM (2009) Molecular mechanisms of heavy metal tolerance and evolution in invertebrates. *Insect Science*, **16**, 3–18.
Jozefczak M, Keunen E, Schat H *et al.* (2014) Differential response of Arabidopsis leaves and roots to cadmium: glutathione-related chelating capacity vs antioxidant capacity. *Plant physiology and biochemistry : PPB*, **83**, 1–9.
Junaid M, Hashmi MZ, Malik RN, Pei D-S (2016) Toxicity and oxidative stress induced by chromium in workers exposed from different occupational settings around the globe: A review. *Environmental science and pollution research international*, **23**, 20151–20167.
Jung D, Adamo MA, Lehman RM *et al.* (2015) A novel variant of aquaporin 3 is expressed in killifish (Fundulus heteroclitus) intestine. *Comparative biochemistry and physiology. Toxicology & pharmacology : CBP*, **171**, 1–7.
Kanehisa M, Goto S (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, **28**, 27–30.
Kanehisa M, Araki M, Goto S *et al.* (2007) KEGG for linking genomes to life and the environment. *Nucleic Acids Research*, **36**, D480–D484.
Kang C-K, Yang W-K, Lin S-T *et al.* (2013) The acute and regulatory phases of time-course changes in gill mitochondrion-rich cells of seawater-acclimated medaka (*Oryzias dancena*) when exposed to hypoosmotic environments. *Comparative Biochemistry and Physiology, Part A*, **164**, 181–191.
Katoh F (2003) Short-term transformation and long-term replacement of branchial chloride cells in killifish transferred from seawater to freshwater, revealed by morphofunctional observations and a newly established `time-differential double fluorescent staining' technique. *Journal of Experimental Biology*, **206**, 4113–4123.
Khaitovich P, Enard W, Lachmann M, Pääbo S (2006) Evolution of primate gene expression. *Nature Reviews Genetics*, **7**, 693–702.
Khaitovich P, Pääbo S, Weiss G (2005) Toward a Neutral Evolutionary Model of Gene Expression. *Genetics*, **170**, 929–939.
Khaitovich P, Weiss G, Lachmann M *et al.* (2004) A Neutral Model of Transcriptome Evolution. *PLoS Biology*, **2**, e132.
Khatri P, Sirota M, Butte AJ (2012) Ten Years of Pathway Analysis: Current Approaches and Outstanding Challenges. *PLoS Computational Biology*, **8**, e1002375.
Ki J-S, Raisuddin S, Lee K-W *et al.* (2009) Gene expression profiling of copper-induced responses in the intertidal copepod *Tigriopus japonicus* using a 6K oligochip microarray. *Aquatic toxicology (Amsterdam, Netherlands)*, **93**, 177–187.
Kilham SS, Kreeger DA, Lynn SG, Goulden CE, Herrera L (1998) COMBO: a defined freshwater culture medium for algae and zooplankton. *Hydrobiologia*, **377**, 147–159.
Kim E, Na KJ (1990) Acute toxic effect of sodium dichromate on metabolism. *Archives of Toxicology*, **64**, 644–649.
Kim HJ, Koedrith P, Seo YR (2015) Ecotoxicogenomic approaches for understanding

molecular mechanisms of environmental chemical toxicity using aquatic invertebrate, *Daphnia* model organism. *International Journal of Molecular Sciences*, **16**, 12261–12287.
Kim J, Lee JJ, Kim J *et al.* (2010) Arsenic antagonizes the Hedgehog pathway by preventing ciliary accumulation and reducing stability of the Gli2 transcriptional effector. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 13432–13437.
Kitano H (2002) Systems Biology: A Brief Overview. *Science*, **295**, 1662–1664.
Klein CA (2013) Selection and adaptation during metastatic cancer progression. *Nature*, **501**, 365–372.
Kozak GM, Brennan RS, Berdan EL, Fuller RC, Whitehead A (2013) Functional and population genomic divergence within and between two species of killifish adapted to different osmotic niches. *Evolution*, **68**, 63–80.
Kozul CD, Ely KH, Enelow RI, Hamilton JW (2009a) Low-Dose Arsenic Compromises the Immune Response to Influenza A Infection in Vivo. *Environmental Health Perspectives*, **117**, 1441–1447.
Kozul CD, Hampton TH, Davey JC *et al.* (2009b) Chronic exposure to arsenic in the drinking water alters the expression of immune response genes in mouse lung. *Environmental Health Perspectives*, **117**, 1108–1115.
Kreeger PK, Lauffenburger DA (2010) Cancer systems biology: a network modeling perspective. *Carcinogenesis*, **31**, 2–8.
Kuhn TS (1976) *The Structure of Scientific Revolutions (Chicago, 1962)*. Kuhn: The Structure of Scientific Revolutions1962.
Kultz D, Avila K (2001) Mitogen-activated protein kinases are in vivo transducers of osmosensory signals in fish gill cells. *Comparative biochemistry and physiology. Part B, Biochemistry & molecular biology*, **129**, 821–829.
Kültz D (2012) The combinatorial nature of osmosensing in fishes. *Physiology (Bethesda, Md.)*, **27**, 259–275.
Kültz D (2015) Physiological mechanisms used by fish to cope with salinity stress. *The Journal of experimental biology*, **218**, 1907–1914.
Kültz D, Li J, Gardell A, Sacchi R (2013) Quantitative molecular phenotyping of gill remodeling in a cichlid fish responding to salinity stress. *Molecular & cellular proteomics : MCP*, **12**, 3962–3975.
Lai KP, Li J-W, Gu J *et al.* (2015) Transcriptomic analysis reveals specific osmoregulatory adaptive responses in gill mitochondria-rich cells and pavement cells of the Japanese eel. *BMC Genomics*, **16**, 1072.
Lam SH, Lui EY, Li Z *et al.* (2014) Differential transcriptomic analyses revealed genes and signaling pathways involved in iono-osmoregulation and cellular remodeling in the gills of euryhaline Mozambique tilapia, *Oreochromis mossambicus*. *BMC Genomics*, **15**, 921.
Leclerc RD (2008) Survival of the sparsest: robust gene networks are parsimonious. *Molecular Systems Biology*, **4**, 213.
Lee TH, Linstedt AD (1999) Osmotically Induced Cell Volume Changes Alter Anterograde and Retrograde Transport, Golgi Structure, and COPI Dissociation. *Molecular Biology of the Cell*, **10**, 1445–1462.
Li M (2003) Cadmium directly induced the opening of membrane permeability pore of mitochondria which possibly involved in cadmium-triggered apoptosis. *Toxicology*, **194**, 19–33.
Li X, Schuler MA, Berenbaum MR (2006) Molecular Mechanisms of Metabolic Resistance to Synthetic and Natural Xenobiotics. *dx.doi.org*, **52**, 231–253.
Liu JT, Bain LJ (2014) Arsenic inhibits hedgehog signaling during P19 cell differentiation.

Toxicology and Applied Pharmacology, **281**, 243–253.
Liu J, Qu W, Kadiiska MB (2009) Role of oxidative stress in cadmium toxicity and carcinogenesis. *Toxicology and Applied Pharmacology*, **238**, 209–214.
Liu S-T, Horng J-L, Chen P-Y, Hwang P-P, Lin L-Y (2016) Salt secretion is linked to acid-base regulation of ionocytes in seawater-acclimated medaka: new insights into the salt-secreting mechanism. *Scientific Reports*, **6**, 1151.
Liu Y, Tennant DA, Zhu Z *et al.* (2014) DiME: A Scalable Disease Module Identification Algorithm with Application to Glioma Progression (RA Calogero, Ed,). **9**, e86693–17.
Lopez J, Colbourne J (2011) Dual-labeled expression microarray protocol for high-throughput genomic investigations. *CGB Technical Report*.
Lu PD, Harding HP, Ron D (2004) Translation reinitiation at alternative open reading frames regulates gene expression in an integrated stress response. *The Journal of Cell Biology*, **167**, 27–33.
Manu, Surkova S, Spirov AV *et al.* (2009) Canalization of Gene Expression and Domain Shifts in the Drosophila Blastoderm by Dynamical Attractors (S Shvartsman, Ed,). *PLoS Computational Biology*, **5**.
Marbach D, Prill RJ, Schaffter T *et al.* (2010) Revealing strengths and weaknesses of methods for gene network inference. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 6286–6291.
Marchetti C (2013) Role of calcium channels in heavy metal toxicity. *ISRN toxicology*.
Marshall WS (2003) Rapid regulation of NaCl secretion by estuarine teleost fish: coping strategies for short-duration freshwater exposures. *Biochimica et biophysica acta*, **1618**, 95–105.
Marshall WS (2011) Mechanosensitive signalling in fish gill and other ion transporting epithelia. *Acta physiologica (Oxford, England)*, **202**, 487–499.
Marshall WS, Cozzi RRF, Pelis RM, McCormick SD (2005) Cortisol receptor blockade and seawater adaptation in the euryhaline teleost *Fundulus heteroclitus*. *Journal of experimental zoology. Part A, Comparative experimental biology*, **303**, 132–142.
Marshall WS, Watters KD, Hovdestad LR, Cozzi RRF, Katoh F (2009) CFTR Cl- channel functional regulation by phosphorylation of focal adhesion kinase at tyrosine 407 in osmosensitive ion transporting mitochondria rich cells of euryhaline killifish. *Journal of Experimental Biology*, **212**, 2365–2377.
Marshall W, Emberley T, Singer T, Bryson S, Mccormick S (1999) Time course of salinity adaptation in a strongly euryhaline estuarine teleost, *Fundulus heteroclitus*: a multivariable approach. *The Journal of experimental biology*, **202 (Pt 11)**, 1535–1544.
Martelli A, Rousselet E, Dycke C, Bouron A, Moulis JM (2006) Cadmium toxicity in animal cells by interference with essential metals. *Biochimie*, **88**, 1807–1814.
Mattingly CJ, Hampton TH, Brothers KM, Griffin NE, Planchart A (2009) Perturbation of defense pathways by low-dose arsenic exposure in zebrafish embryos. *Environmental Health Perspectives*, **117**, 981–987.
Mähler N, Wang J, Terebieniec BK *et al.* (2017) Gene co-expression network connectivity is an important determinant of selective constraint. (NM Springer, Ed,). *PLoS Genetics*, **13**, e1006402.
Michalak P (2008) Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes. *Genomics*, **91**, 243–248.
Mizushima N, Komatsu M (2011) Autophagy: Renovation of Cells and Tissues. *Cell*, **147**, 728–741.
Moorman BP, Lerner DT, Grau EG, Seale AP (2015) The effects of acute salinity challenges on osmoregulation in Mozambique tilapia reared in a tidally changing salinity. *The Journal of experimental biology*, **218**, 731–739.

Muyssen BTA, Janssen CR (2004) Multi-generation cadmium acclimation and tolerance in Daphnia magna Straus. *Environmental pollution (Barking, Essex : 1987)*, **130**, 309–316.
Naujokas MF, Anderson B, Ahsan H *et al.* (2013) The Broad Scope of Health Effects from Chronic Arsenic Exposure: Update on a Worldwide Public Health Problem. *Environmental Health Perspectives*, **121**, 295–302.
Necsulea A, Kaessmann H (2014) Evolutionary dynamics of coding and non-coding transcriptomes. *Nature Reviews Genetics*, **15**, 734–748.
Nguyen AH, Altomare LE, McElwain MC (2014) Decreased accumulation of cadmium in Drosophila selected for resistance suggests a mechanism independent of metallothionein. *Biological trace element research*, **160**, 245–249.
Nota B, de Korte M, Ylstra B, van Straalen NM, Roelofs D (2013) Genetic Variation in Parthenogenetic Collembolans Is Associated with Differences in Fitness and Cadmium-Induced Transcriptome Responses. *Environmental Science & Technology*, **47**, 1155–1162.
Oerton E, Bender A (2017) Concordance analysis of microarray studies identifies representative gene expression changes in Parkinson's disease: a comparison of 33 human and animal studies. *BMC Neurology*, **17**, 58.
Oleksiak MF, Churchill GA, Crawford DL (2002) Variation in gene expression within and among natural populations. *Nature Genetics*, **32**, 261–266.
Oleksiak MF, Roach JL, Crawford DL (2005) Natural variation in cardiac metabolism and gene expression in *Fundulus heteroclitus*. *Nature Genetics*, **37**, 67–72.
Olsen CE, Liguori AE, Zong Y *et al.* (2008) Arsenic upregulates MMP-9 and inhibits wound repair in human airway epithelial cells. *AJP: Lung Cellular and Molecular Physiology*, **295**, L293–302.
Owen J, Hedley BA, Svendsen C *et al.* (2008) Transcriptome profiling of developmental and xenobiotic responses in a keystone soil animal, the oligochaete annelid *Lumbricus rubellus*. *BMC Genomics*, **9**, 266.
Palumbi SR (2001) Humans as the World's Greatest Evolutionary Force. *Science*, **293**, 1786–1790.
Pang CNI, Lai Y-W, Campbell LT *et al.* (2017) Transcriptome and network analyses in Saccharomyces cerevisiae reveal that amphotericin B and lactoferrin synergy disrupt metal homeostasis and stress response. *Scientific Reports*, **7**, 40232.
Paoliello MMB, De Capitani EM, da Cunha FG *et al.* (2002) Exposure of Children to Lead and Cadmium from a Mining Area of Brazil. *Environmental Research*, **88**, 120–128.
Paul DS, Harmon AW, Devesa V, Thomas DJ, Styblo M (2007) Molecular Mechanisms of the Diabetogenic Effects of Arsenic: Inhibition of Insulin Signaling by Arsenite and Methylarsonous Acid. *Environmental Health Perspectives*, **115**, 734–742.
Pedrosa J, Gravato C, Campos D *et al.* (2017) Investigating heritability of cadmium tolerance in *Chironomus riparius* natural populations: A physiological approach. *Chemosphere*, **170**, 83–94.
Phillips ND (2017) YaRrr! The Pirate's Guide to R. *APS Observer*, **30**.
Pinheiro J, Bates D, DebRoy S, Sarkar D, Team RC (2015) *nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1–120 http*: CRAN.R-project.org/
Platt JR (1964) Strong Inference: Certain systematic methods of scientific thinking may produce much more rapid progress than others. *Science*, **146**, 347–353.
Popper KR (2002) The Logic of Scientific Discovery (Routledge Classics). *(31 March 2002)*.
Poynton, Julia R Varshavsky, Bonnie Chang *et al.* (2006) *Daphnia magna* Ecotoxicogenomics Provides Mechanistic Insights into Metal Toxicity. *Environmental Science & Technology*, **41**, 1044–1050.
Prévéral S, Gayet L, Moldes C *et al.* (2009) A Common Highly Conserved Cadmium

Detoxification Mechanism from Bacteria to Humans HEAVY METAL TOLERANCE CONFERRED BY THE ATP-BINDING CASSETTE (ABC) TRANSPORTER SpHMT1 REQUIRES GLUTATHIONE BUT NOT METAL-CHELATING PHYTOCHELATIN PEPTIDES. *Journal of Biological Chemistry*, **284**, 4936–4943.
Pytharopoulou S, Grintzalis K, Sazakli E *et al.* (2011) Translational responses and oxidative stress of mussels experimentally exposed to Hg, Cu and Cd: One pattern does not fit at all. *Aquatic Toxicology*, **105**, 157–165.
Rand du EE, Smit S, Beukes M *et al.* (2015) Detoxification mechanisms of honey bees (*Apis mellifera*) resulting in tolerance of dietary nicotine. *Scientific Reports*, **5**, 11779.
Ranta E, Bengtsson J, McManus J (1993) Accept Terms and Conditions on JSTOR. *Annales Zoologici Fennici*.
Ratnaike RN (2003) Acute and chronic arsenic toxicity. *Postgraduate medical journal*, **79**, 391–396.
Reid NM, Jackson CE, Gilbert D *et al.* (2017) The landscape of extreme genomic variation in the highly adaptable Atlantic killifish. *Genome Biology and Evolution*.
Reid NM, Proestou DA, Clark BW *et al.* (2016) The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish. *Science*, **354**, 1305–1308.
Revelle W (2014) psych: Procedures for psychological, psychometric, and personality research. *Northwestern University*.
Roelofs D, Janssens TKS, Timmermans MJTN *et al.* (2009) Adaptive differences in gene expression associated with heavy metal tolerance in the soil arthropod *Orchesella cincta*. *Molecular Ecology*, **18**, 3227–3239.
Roelofs D, Morgan J, Stürzenbaum S (2010) The significance of genome-wide transcriptional regulation in the evolution of stress tolerance. *Evolutionary Ecology*, **24**, 527–539.
Rosen BP, Ajees AA, McDermott TR (2011) Life and death with arsenic. Arsenic life: an analysis of the recent report "A bacterium that can grow by using arsenic instead of phosphorus". *BioEssays : news and reviews in molecular, cellular and developmental biology*, **33**, 350–357.
Rubino F (2015) Toxicity of Glutathione-Binding Metals: A Review of Targets and Mechanisms. *Toxics*, **3**, 20–62.
Sanit di Toppi L, Gabbrielli R (1999) Response to cadmium in higher plants. *Environmental and experimental botany*, **41**, 105–130.
Sanjaya, Hsiao P-Y, Su R-C *et al.* (2008) Overexpression of *Arabidopsis thaliana* tryptophan synthase beta 1 (AtTSB1) in Arabidopsis and tomato confers tolerance to cadmium stress. *Plant, cell & environment*, **31**, 1074–1085.
Sarnelle O, Wilson AE (2005) Local adaptation of *Daphnia pulicaria* to toxic cyanobacteria. *Limnology and Oceanography*, **50**, 1565–1570.
Satarug S, Vesey DA, Gobe GC (2017a) Health Risk Assessment of Dietary Cadmium Intake: Do Current Guidelines Indicate How Much is Safe? *Environmental Health Perspectives*, **125**, 284–288.
Satarug S, Vesey DA, Gobe GC (2017b) Current health risk assessment practice for dietary cadmium: Data from different countries. *Food and Chemical Toxicology*, **106**, 430–445.
Schulte PM (2007) Responses to environmental stressors in an estuarine fish: Interacting stressors and the impacts of local adaptation. *Journal of Thermal Biology*, **32**, 152–161.
Seale AP, Watanabe S, Breves JP *et al.* (2012) Differential regulation of TRPV4 mRNA levels by acclimation salinity and extracellular osmolality in euryhaline tilapia. *General and Comparative Endocrinology*, **178**, 123–130.
Segal E, Shapira M, Regev A *et al.* (2003) Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nature Genetics*, **34**, 166–176.

Shaw JR, Bomberger JM, VanderHeide J *et al.* (2010) Arsenic inhibits SGK1 activation of CFTR Cl- channels in the gill of killifish, *Fundulus heteroclitus*. *Aquatic toxicology (Amsterdam, Netherlands)*, **98**, 157–164.
Shaw JR, Colbourne JK, Davey JC *et al.* (2007) Gene response profiles for *Daphnia pulex* exposed to the environmental stressor cadmium reveals novel crustacean metallothioneins. *BMC Genomics*, **8**, 477–20.
Shaw JR, Dempsey TD, Chen CY, Hamilton JW, Folt CL (2006) Comparative toxicity of cadmium, zinc, and mixtures of cadmium and zinc to daphnids. *Environmental toxicology and chemistry / SETAC*, **25**, 182–189.
Shaw JR, Hampton TH, King BL *et al.* (2014) Natural selection canalizes expression variation of environmentally induced plasticity-enabling genes. *Molecular Biology and Evolution*, **31**, 3002–3015.
Shaw JR, Pfrender ME, Eads BD *et al.* (2008) Daphnia as an emerging model for toxicological genomics. In: *Comparative Toxicogenomics* Advances in Experimental Biology. pp. 165–328. Elsevier.
Shi L, Reid LH, Jones WD *et al.* (2006) The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nature Biotechnology*, **24**, 1151–1161.
Shim J, Kennedy RH, Weatherly LM *et al.* (2016) Arsenic inhibits mast cell degranulation via suppression of early tyrosine phosphorylation events. *Journal of Applied Toxicology*, **36**, 1446–1459.
Shinkai Y, Sumi D, Toyama T, Kaji T, Kumagai Y (2009) Role of aquaporin 9 in cellular accumulation of arsenic and its cytotoxicity in primary mouse hepatocytes. *Toxicology and Applied Pharmacology*, **237**, 232–236.
Sinha M, Manna P, Sil PC (2008) Cadmium-induced neurological disorders: prophylactic role of taurine. *Journal of Applied Toxicology*, **28**, 974–986.
Smith AH, Marshall G, Yuan Y *et al.* (2006) Increased Mortality from Lung Cancer and Bronchiectasis in Young Adults after Exposure to Arsenic in Utero and in Early Childhood. **114**, 1293–1296.
Soetaert A, Vandenbrouck T, van der Ven K *et al.* (2007) Molecular responses during cadmium-induced stress in Daphnia magna: Integration of differential gene expression with higher-level effects. *Aquatic Toxicology*, **83**, 212–222.
Stanton CR, Thibodeau R, Lankowski A *et al.* (2006) Arsenic inhibits CFTR-mediated chloride secretion by killifish (*Fundulus heteroclitus*) opercular membrane. *Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology*, **17**, 269–278.
Stephen R Stürzenbaum, Oleg Georgiev, A John Morgan A, Kille P (2004) Cadmium Detoxification in Earthworms: From Genes to Cells†. *Environmental Science & Technology*.
Stincone A, Prigione A, Cramer T *et al.* (2015) The return of metabolism: biochemistry and physiology of the pentose phosphate pathway. *Biological Reviews*, **90**, 927–963.
Stuhlbacher A, Bradley MC, Naylor C, Calow P (1992) Induction of cadmium tolerance in two clones of Daphnia magna straus. *Comparative Biochemistry and Physiology Part C: Comparative Pharmacology*, **101**, 571–577.
Svendsen C, Owen J, Kille P *et al.* (2008) Comparative transcriptomic responses to chronic cadmium, fluoranthene, and atrazine exposure in Lumbricus rubellus. *Environmental Science & Technology*, **42**, 4208–4214.
Tawfik DS, Viola RE (2011) Arsenate Replacing Phosphate: Alternative Life Chemistries and Ion Promiscuity. *Biochemistry*, **50**, 1128–1134.
Team RC (2014a) *R: A Language and Environment for Statistical Computing, version 3.1. 0.*

Vienna: R Foundation for Statistical Computing. Available at http://www. R-project. org.
Team RC (2014b) *R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria*.
Timmermans MJTN, Ellers J, Roelofs D, van Straalen NM (2005) Metallothionein mRNA Expression and Cadmium Tolerance in Metal-stressed and Reference Populations of the Springtail *Orchesella cincta*. *Ecotoxicology (London, England)*, **14**, 727–739.
Tsapakos MJ, Hampton TH, Jennette KW (1981) The carcinogen chromate induces DNA cross-links in rat liver and kidney. *Journal of Biological Chemistry*, **256**, 3623–3626.
Tsapakos MJ, Hampton TH, Wetterhahn KE (1983a) Chromium(VI)-induced DNA lesions and chromium distribution in rat kidney, liver, and lung. *Cancer Research*, **43**, 5662–5667.
Tsapakos MJ, Hampton TH, Sinclair PR *et al.* (1983b) The carcinogen chromate causes DNA damage and inhibits drug-mediated induction of porphyrin accumulation and glucuronidation in chick embryo hepatocytes. *Carcinogenesis*, **4**, 959–966.
Tukey JW (1977) *Exploratory Data Analysis*. Pearson College Division.
Uebbing S, Künstner A, Mäkinen H *et al.* (2016) Divergence in gene expression within and between two closely related flycatcher species. *Molecular Ecology*, **25**, 2015–2028.
van Doorslaer W, Vanoverbeke J, Duvivier C *et al.* (2009) Local adaptation to higher temperatures reduces immigration success of genotypes from a warmer region in the water flea Daphnia. *Global Change Biology*, **15**, 3046–3055.
Van Straalen NM, Roelofs D (2005) Cadmium tolerance in a soil arthropod. *Entomologische Berichten*. **65**(4): 105-111
van Straalen NM, Janssens TKS, Roelofs D (2011) Micro-evolution of toxicant tolerance: from single genes to the genome's tangled bank. *Ecotoxicology (London, England)*, **20**, 574–579.
Velotta JP, Wegrzyn JL, Ginzburg S *et al.* (2017) Transcriptomic imprints of adaptation to fresh water: parallel evolution of osmoregulatory gene expression in the Alewife. *Molecular Ecology*, **26**, 831–848.
Vogelstein B, Kinzler KW (2004) Cancer genes and the pathways they control. *Nature Medicine*, **10**, 789–799.
Wang X, Meng D, Chang Q *et al.* (2010) Arsenic inhibits neurite outgrowth by inhibiting the LKB1-AMPK signaling pathway. *Environmental Health Perspectives*, **118**, 627–634.
Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, **10**, 57–63.
Ward TJ, Robinson WE (2005) Evolution of cadmium resistance in *Daphnia magna*. *Environmental Toxicology and Chemistry*, **24**, 2341–2349.
Wen G, Hong M, Calaf GM, Roy D (2010) Phosphoproteomic profiling of arsenite-treated human small airway epithelial cells. *Oncologyl Reports,* **23**(2):405-12.
Wetzler M, Brady MT, Tracy E *et al.* (2006) Arsenic Trioxide Affects Signal Transducer and Activator of Transcription Proteins through Alteration of Protein Tyrosine Kinase Phosphorylation. *Clinical Cancer Research*, **12**, 6817–6825.
Whitehead A (2009) Comparative mitochondrial genomics within and among species of killifish. *BMC evolutionary biology*, **9**, 11.
Whitehead A (2012) Comparative genomics in ecological physiology: toward a more nuanced understanding of acclimation and adaptation. *Journal of Experimental Biology*, **215**, 884–891.
Whitehead A (2014) Evolutionary Genomics of Environmental Pollution. In: *Ecological Genomics* Advances in Experimental Medicine and Biology. pp. 321–337. Springer Netherlands, Dordrecht.
Whitehead A, Crawford DL (2006) Neutral and adaptive variation in gene expression.

Proceedings of the National Academy of Sciences, **103**, 5425–5430.
Whitehead A, Galvez F, Zhang S, Williams LM, Oleksiak MF (2011a) Functional genomics of physiological plasticity and local adaptation in killifish. *The Journal of heredity*, **102**, 499–511.
Whitehead A, Pilcher W, Champlin D, Nacci D (2012a) Common mechanism underlies repeated evolution of extreme pollution tolerance. *Proceedings of the Royal Society B: Biological Sciences*, **279**, 427–433.
Whitehead A, Roach JL, Zhang S, Galvez F (2011b) Genomic mechanisms of evolved physiological plasticity in killifish distributed along an environmental salinity gradient. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 6193–6198.
Whitehead A, Roach JL, Zhang S, Galvez F (2012b) Salinity- and population-dependent genome regulatory response during osmotic acclimation in the killifish (*Fundulus heteroclitus*) gill. *Journal of Experimental Biology*, **215**, 1293–1305.
Whitehead A, Triant DA, Champlin D, Nacci D (2010) Comparative transcriptomics implicates mechanisms of evolved pollution tolerance in a killifish population. *Molecular Ecology*, **19**, 5186–5203.
Whitehead A, Zhang S, Roach JL, Galvez F (2013) Common functional targets of adaptive micro- and macro-evolutionary divergence in killifish. *Molecular Ecology*, **22**, 3780–3796.
Wirgin I, Roy NK, Loftus M *et al.* (2011) Mechanistic basis of resistance to PCBs in Atlantic tomcod from the Hudson River. *Science*, **331**, 1322–1325.
Xu S, He M, Zhong M *et al.* (2015) The neuroprotective effects of taurine against nickel by reducing oxidative stress and maintaining mitochondrial function in cortical neurons. *Neuroscience Letters*, **590**, 52–57.
Yamaguchi C, Ohkama-Ohtsu N, Shinano T, Maruyama-Nakashita A (2017) Plants prioritize phytochelatin synthesis during cadmium exposure even under reduced sulfate uptake caused by the disruption of SULTR1;2. *Plant Signaling & Behavior*, **12**, e1325053.
Yamasaki S, Anderson P (2008) Reprogramming mRNA translation during stress. *Current opinion in cell biology*, **20**, 222–226.
Yip AM, Horvath S (2007) BMC Bioinformatics. **8**, 22–14.
Zhang H-N, Yang L, Ling J-Y *et al.* (2015) Systematic identification of arsenic-binding proteins reveals that hexokinase-2 is inhibited by arsenic. *Proceedings of the National Academy of Sciences*, **112**, 15084–15089.
Zhang M, Yao C, Guo Z *et al.* (2008) Apparently low reproducibility of true differential expression discoveries in microarray studies. *Bioinformatics*, **24**, 2057–2063.